

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:00:47 : Search time 45 Seconds
(without alignments)
704,748 Million cell updates/sec

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELFTGVVILVELDG.....VLLEFVTAAGITHCMDELYK 238

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1275	100.0	238	17 AAW05304	Green fluorescent
2	1275	100.0	238	18 AAW24232	Aequorea victoria
3	1275	100.0	238	19 AAW76371	A. victoria green
4	1275	100.0	238	19 AAW76105	A. victoria green
5	1275	100.0	238	19 AAW40479	A. victoria green
6	1275	100.0	238	22 AAB73552	Wild-type green fl
7	1275	100.0	238	23 AAE16038	Aequorea victoria
8	1274	99.9	238	19 AAW65084	A. victoria green
9	1272	99.8	238	19 AAW76109	A. victoria green
10	1272	99.8	238	19 AAW76110	A. victoria green

11	1272	99.8	238	19 AAW76111	A. victoria green
12	1272	99.8	238	19 AAW5079	A. victoria green
13	1272	99.8	238	19 AAW5079	A. victoria green
14	1272	99.8	238	23 AAE16042	Aequorea victoria
15	1272	99.8	238	23 AAE16042	Aequorea victoria
16	1272	99.8	501	18 AAW31879	GFP variants S65T
17	1271	99.7	238	17 AAW05312	Green fluorescent
18	1271	99.7	238	18 AAW36057	Green fluorescent
19	1271	99.7	238	18 AAW31295	Aequorea victoria
20	1271	99.7	238	19 AAW76113	A. victoria green
21	1271	99.7	238	19 AAW65081	A. victoria green
22	1271	99.7	238	19 AAW65085	A. victoria green
23	1271	99.7	238	20 AAY43249	Green fluorescent
24	1271	99.7	238	20 AAW88484	Green fluorescent
25	1271	99.7	238	21 AAB01616	Aequorea victoria
26	1271	99.7	238	21 AAB03411	Aequorea victoria
27	1271	99.7	238	21 AAB15661	Humanised GFP, Ho
28	1271	99.7	238	21 AAB01658	Aequorea victoria
29	1271	99.7	238	21 AAY99804	Aequorea victoria
30	1271	99.7	238	21 AAY99835	Aequorea victoria
31	1271	99.7	238	21 AAY99723	Aequorea victoria
32	1271	99.7	238	21 AAY99732	Aequorea victoria
33	1271	99.7	238	21 AAY70365	Humanised green fl
34	1271	99.7	238	22 AAB73554	Green fluorescent
35	1271	99.7	238	22 AAB85899	A. victoria green
36	1271	99.7	238	22 AAG66199	A. victoria green
37	1271	99.7	238	23 AAE14600	Aequorea victoria
38	1271	99.7	238	23 AAE16041	Aequorea victoria
39	1271	99.7	238	23 AAE16056	Aequorea victoria
40	1271	99.7	239	21 AAY59283	Amino acid sequenc
41	1270	99.6	238	17 AAW05306	Green fluorescent
42	1270	99.6	238	19 AAW59822	Green fluorescent
43	1270	99.6	238	19 AAW65086	A. victoria green
44	1270	99.6	238	19 AAW65082	A. victoria green
45	1270	99.6	238	22 AAB73555	Green fluorescent

ALIGNMENTS

RESULT 1

AAW05304

ID AAW05304 standard; Protein: 238 AA.

XX AAW05304:

AC AAW05304:

DT 02-APR-1997 (first entry)

XX Green fluorescent protein.

DE Green fluorescent protein.

XX Green fluorescent protein: GFP; jellyfish; Aequorea victoria; sea pansy;

KW Renilla reniformis; differential gene expression; protein localisation;

KW gene expression tracking; fluorescence.

XX Aequorea victoria.

OS Aequorea victoria.

PN WO9623810-A1

XX 08-AUG-1996.

PD 08-AUG-1996.

XX 13-NOV-1995: 95WO-US14692.

PF 10-NOV-1994; 94US-037915.

PR (REGC) UNIV CALIFORNIA.

XX Heim R. Tsien RV;

XX WPI: 1996-371370/37.

DR N-PSDB: AAT39694.

XX New modified Aequorea green fluorescent polypeptide(s) - having

PT amino acid changes to provide prods. which exhibit different

PT excitation and emission spectra
 XX
 PS
 XX Disclosure; Page 20-21; 39pp; English.
 XX
 CC This sequence represents the wild type Aequorea victoria (jellyfish)
 CC green fluorescent protein (GFP). The fluorescence of this protein is
 CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
 CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
 CC the one absorbance peak seen in the related GFP from the sea pansy
 CC (Renilla reniformis). This protein is modified, to provide the GFPs of
 CC the invention. The modifications were created by subjecting the cDNA
 CC encoding this sequence to site directed mutagenesis using mutagenic PCR
 CC primers, or hydroxylamine treatment. The GFPs of the invention are
 CC modified to lead to the formation of products with markedly different
 CC excitation and emission spectra. Visibly distinct colours, and increased
 CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.
 XX
 SQ Sequence 238 AA;
 Query Match 100.0%; Score 1275; PB 17; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIILVELDGVNGHKFVSVSSEGGDGYGKLTFLKFICTTGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFVSVSSEGGDGYGKLTFLKFICTTGKLPVNPPTL 60
 QY 61 VTFYSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTFYSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180
 DB 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGTTHGMDELYK 238
 DB 161 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGTTHGMDELYK 238
 RESULT 2
 AAW24232
 ID AAW24232 standard; Protein; 238 AA.
 XX
 AC AAW24232;
 XX
 DT 02-FEB-1998 (first entry)
 XX
 DE Aequorea victoria green fluorescent protein (GFP).
 XX
 KW North West Pacific jellyfish; green fluorescent protein; GFP; S65C;
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; Tandem fluorescent protein construct;
 KW blue fluorescent protein; protease cleavage site.
 XX
 OS Aequorea victoria.
 XX
 PN WO9728261-A1.
 XX
 PD 07-AUG-1997.
 XX
 PF 31-JAN-1997; 97WO-US01457.
 XX
 XX 31-JAN-1996; 96US-0594575.
 PR
 XX (AURO-) ACURA BIOSCIENCES CORP.
 PA (REGC) UNIV CALIFORNIA.
 XX
 XX Cubitt A, Heim R, Tsien RY;
 XX

DR WPI; 1997-402615/37.
 DR N-PSDR; AAT77843.
 XX
 PT Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker.
 PT useful in enzymatic assays
 XX
 PS Disclosure; Fig 1; 88pp; English.
 XX
 CC This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of
 CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.
 XX
 SQ Sequence 238 AA;
 Query Match 100.0%; Score 1275; DB 18; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVPIILVELDGVNGHKFVSVSSEGGDGYGKLTFLKFICTTGKLPVNPPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFVSVSSEGGDGYGKLTFLKFICTTGKLPVNPPTL 60
 QY 61 VTFYSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTFYSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180
 DB 121 NRTELKGIQDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGTTHGMDELYK 238
 RESULT 3
 AAW76371
 ID AAW76371 standard; Protein; 238 AA.
 XX
 AC AAW76371;
 XX
 DT 03-DEC-1998 (first entry)
 XX
 DE A. victoria green fluorescent protein.
 XX
 KW Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.
 XX
 OS Aequorea victoria.
 XX
 PN JP10234382-A.
 XX
 PD 02-SEP-1998.
 XX
 PF 27-FEB-1997; 97JP-0062370.
 XX
 PR 27-FEB-1997; 97JP-0062370.
 XX
 XX (DINA-) DINABEKU KENYUSHO KK.
 PA
 XX

RESULT 3
AAE16073
ID AAE16073 standard; Protein: 238 AA.
XX AC AAE16073;
XX
DT 26-MAR-2002 (first entry)
XX
DE Aequorea victoria GFP mutant protein (E222X).
XX
KW Green fluorescent protein: cell lineage tracer; protein localisation:
KW GFP: fusion tag; gene expression marker; fluorescent energy acceptor;
KW immunocassay: hybridisation assay; fluorescent energy donor; biosensor;
KW FRET; fluorescence resonance energy transfer; mutant; mutein.
XX
OS Aequorea victoria.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 222 /label= Asn, Gln
FT /note= "Wild type Glu substituted with Xaa"
FT
XX WC200190147-A2.
XX
PD 29-NOV-2001.
XX
PF 17-MAY-2001: 2001WO-US16149.
XX
PR 19-MAY-2000: 2000US-0575847.
XX
FA (UYOR-) UNIV OREGON STATE.
XX
PI Wachter R, Remington SJ;
XX
CR WPI: 2002-083084/11.
XX
PT New long wavelength engineered fluorescent proteins, useful as markers
PT for gene expression, tracers of cell lineage or as fusion tags to
PT monitor protein localization, or in detection assays, e.g. immunoassays
PT or hybridization assays .
XX
PS Claim 47; Page -: 181pp; English.
XX
CC The patent discloses functional engineered fluorescent protein and its
CC corresponding polynucleotide. The amino acid sequence of the engineered
CC protein is identical to Aequorea green fluorescent protein (GFP). The
CC engineered fluorescent proteins of the invention have varied fluorescent
CC properties and have the ability to respond to ion concentrations via a
CC change in fluorescent characteristics. They are useful as markers for
CC gene expression, tracers of cell lineage or as fusion tags to monitor
CC protein localisation within living cells. The engineered fluorescent
CC proteins are particularly useful for coupling engineered fluorescent
CC proteins to antibodies, nucleic acids or other receptors for use in
CC detection assays, e.g. immunoassays or hybridisation assays. They are
CC useful for tracking the movement of proteins in cells or in systems
CC for detecting induction of transcription and for the simultaneous
CC measurement of two or more processes within cells. Proteins of the
CC invention are also useful as fluorescent energy donors or acceptors
CC as well as biosensors for detecting anions. They are also useful in
CC fluorescence resonance energy transfer (FRET). The crystal structure
CC of the GFP is useful for designing mutants having altered fluorescent
CC characteristics which are particularly useful to identify amino acids
CC whose substitution alters fluorescent properties of the protein. The
CC crystal structure of the GFP is also useful for designing mutants having
CC altered anion binding characteristics which are particularly useful for
CC identifying amino acids whose substitution alters the specificity and
CC affinity of the binding site to various anions, and for monitoring anion
CC binding and therefore the concentration of the anion. The present
CC sequence is Aequorea victoria GFP mutant protein (E222X).
CC Note: This sequence is not shown in the specification but is derived
CC from Aequorea victoria wild type green fluorescent protein shown in

CC figure 3 of the specification (AAE16073).
XX
SQ Sequence 238 AA;
Query Match 98.7%; Score 1259; DB 23; Length 238;
Best Local Similarity 98.7%; Pred. No. 1.5e-123;
Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSKGEELFGVWPILVELDGVNGHFKFSVSGEGDATYKGLTLKFLICITGKLPVWPPTL 60
DB 1 MSKGEELFGVWPILVELDGVNGHFKFSVSGEGDATYKGLTLKFLICITGKLPVWPPTL 60
QY 61 VTLSYGVQCFSRYPDBMKRHDFFKSAPEGVYQEMTIFFKDGNKYKTRAEYKFGDTLY 120
DB 61 VTLSYGVQCFSRYPDBMKRHDFFKSAPEGVYQEMTIFFKDGNKYKTRAEYKFGDTLY 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNEKIRHNIEDGSGVLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYIMADKQKNGIKVNEKIRHNIEDGSGVLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMYLLGFVTAAGITHGMDELYK 238
RESULT 4
AAW05304
ID AAW05304 standard; Protein: 238 AA.
XX AC AAW05304;
XX
DT 02-APR-1997 (first entry)
XX
DE Green fluorescent protein.
XX
KW Green fluorescent protein: GFP; jellyfish; Aequorea victoria; sea pansy;
KW Renilla reniformis; differential gene expression; protein localisation;
KW gene expression tracking; fluorescence.
XX
OS Aequorea victoria.
XX
PN WC9623810-A1.
XX
PD 08-AUG-1996.
XX
PF 13-NOV-1995: 95WO-US14692.
XX
PR 10-NOV-1994: 94US-Q337915.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Heim R, Tsien RY;
XX
WPI: 1996-371370/37.
XX N-PSDB: AAI39694.
XX
PT New modified Aequorea green fluorescent polypeptide(s) - having
PT amino acid changes to provide prods. which exhibit different
PT excitation and emission spectra
XX
PS Disclosure: Page 20-21; 39pp; English.
XX
CC This sequence represents the wild type Aequorea victoria (jellyfish)
CC green fluorescent protein (GFP). The fluorescence of this protein is
CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
CC the one absorbance peak seen in the related GFP from the sea pansy
CC (Renilla reniformis). This protein is modified, to provide the GFPs of
CC the invention. The modifications were created by subjecting the cDNA
CC encoding this sequence to site directed mutagenesis using mutagenic PCR
CC primers, or hydroxylamine treatment. The GFPs of the invention are
CC modified to lead to the formation of products with markedly different
CC excitation and emission spectra. Visibly distinct colours, and increased

CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.

XX
 SQ Sequence 238 AA;
 Query Match 98.6%; Score 1258; DB 17; Length 238;
 Best Local Similarity 98.7%; Pred. No. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSKGEELFTGVVPIVLVDGVDNGHGFVSVEGEGDATYGLIKLFICTTCKLPVPMPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHGFVSVEGEGDATYGLIKLFICTTCKLPVPMPTL 60
 OY 61 VTLSYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTLSYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 OY 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGGVQLAD 180
 DB 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGGVQLAD 180
 OY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDRHMLVLCFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDRHMLVLCFVTAAGITHGMDELYK 238

RESULT 5
 ANW24232
 ID AAW24232 standard; Protein: 238 AA.
 XX
 AC AAW24232;
 DI 02-FEB-1998 (first entry)
 DE Aequorea victoria green fluorescent protein (GFP).
 KW North West Pacific jellyfish; green fluorescent protein; GFP; S65C;
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; Tandem fluorescent protein construct;
 KW blue fluorescent protein; protease cleavage site.
 XX
 OS Aequorea victoria.
 XX
 PN WO9728261-A1.
 XX
 PD 07-AUG-1997.
 XX
 PF 31-JAN-1997; 97WO-US01457.
 XX
 PR 31-JAN-1996; 96US-0594575.
 XX
 PA (AURO-) AURORA BIOSCIENCES CORP.
 XX (REGC) UNIV CALIFORNIA.
 XX
 PI Cubitt A. Heim R. Tsien RY;
 XX
 PP 1997-402615/37.
 DR N-PSDB; AAI77843.
 XX
 PT Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker,
 PT useful in enzymatic assays
 XX
 PS Disclosure; Fig 1; 80pp; English.

XX
 CC This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of

CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties inhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.

XX
 SQ Sequence 238 AA;
 Query Match 98.6%; Score 1258; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. No. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSKGEELFTGVVPIVLVDGVDNGHGFVSVEGEGDATYGLIKLFICTTCKLPVPMPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHGFVSVEGEGDATYGLIKLFICTTCKLPVPMPTL 60
 OY 61 VTLSYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTLSYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
 OY 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGGVQLAD 180
 DB 121 NRTELKGDIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKVNFKIRHNIEDGGVQLAD 180
 OY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDRHMLVLCFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDRHMLVLCFVTAAGITHGMDELYK 238

RESULT 6
 AAW76371
 ID AAW76371 standard; Protein: 238 AA.
 XX
 AC AAW76371;
 DI 03-DEC-1998 (first entry)
 DE A. victoria green fluorescent protein.
 KW Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.
 XX
 OS Aequorea victoria.
 XX
 PN JPI0234382-A.
 XX
 PD 08-SEP-1998.
 XX
 PF 27-FEB-1997; 97JP-0062370.
 XX
 PR 27-FEB-1997; 97JP-0062370.
 XX
 PA (JINA-) DINABEKKU KENYUSHO KK.
 XX
 PP WPI; 1998-535036/46.
 DR N-PSDB; AAV61751.
 XX
 PT Green fluorescent protein - useful for determining promoter and
 PT targeting activity in a cell
 XX
 PS Disclosure; Page 6-7; 11pp; Japanese.

XX
 CC This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from Aequorea victoria. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 6 times
 CC higher than the conventional fluorescent proteins. The nucleic acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested,
 CC comprising steps of introducing the above vector to a host cell; and

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.333 Seconds

(without alignments)
1196.433 Million cell updates/sec

Title: US-09-967-301-3

Perfect score: 1276

Sequence: MSKGEELFTGVVPIVLVDG.....VLISFVTAAGIHGMDELYK 238

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 471540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-misc:*
- 8: SP-organelle:*
- 9: SP-phage:*
- 10: SP-plant:*
- 11: SP-todent:*
- 12: SP-virus:*
- 13: SP-vertebrate:*
- 14: SP-unclassified:*
- 15: SP-virus:*
- 16: SP-bacteriap:*
- 17: SP-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1254	96.3	238	5 Q27903	Q27903 unidentified
2	1247	97.7	238	5 Q31225	Q31225 aquorea vi
3	1212	95.0	238	5 Q17105	Q17105 aquorea vi
4	1204	94.4	758	4 Q56J01	Q56J01 homo sapien
5	1197	93.8	238	5 Q17106	Q17106 aquorea vi
6	1105	86.6	238	5 Q8WTC6	Q8WTC6 aquorea ma
7	1101	86.3	238	5 Q8WP95	Q8WP95 aquorea ma
8	1097	86.0	238	5 Q8WTC4	Q8WTC4 aquorea ma
9	1095	85.8	238	5 Q8WID0	Q8WID0 aquorea ma
10	1094	85.7	238	5 Q8WTC9	Q8WTC9 aquorea ma
11	1094	85.7	238	5 Q8WTC8	Q8WTC8 aquorea ma
12	1092	85.6	238	5 Q8WTC7	Q8WTC7 aquorea ma
13	1090	85.4	238	5 Q8WTC5	Q8WTC5 aquorea ma
14	252.5	19.8	225	5 Q95UA7	Q95UA7 montastraea
15	250.5	19.6	236	5 Q8T60C	Q8T60C dendronephr
16	250	19.6	225	5 Q563F5	Q563F5 montastraea

17	244.5	19.2	266	5 Q9U6Y3	Q9U6Y3 Clavularia
18	237.5	18.6	225	5 Q8T5F1	Q8T5F1 montastraea
19	211.5	16.6	234	5 Q8T5F2	Q8T5F2 montastraea
20	210.5	16.5	229	5 Q9U6Y5	Q9U6Y5 anemonia ma
21	210	16.5	221	5 Q95P04	Q95P04 gonioporta t
22	209	16.4	227	5 Q962P9	Q962P9 montastraea
23	207.5	16.3	225	5 Q9T6Y8	Q9T6Y8 discosoma s
24	205.5	16.1	232	5 Q9GP15	Q9GP15 anemonia su
25	205	16.1	227	5 Q95V10	Q95V10 montastraea
26	205	16.1	236	5 Q9BLV9	Q9BLV9 renilla muc
27	203.5	15.9	232	5 Q9U6Y7	Q9U6Y7 discosoma s
28	201.5	15.8	232	5 Q9GZ28	Q9GZ28 anemonia su
29	199.5	15.6	235	5 Q8T5F0	Q8T5F0 scolyxia cu
30	197.5	15.5	225	5 Q8T6T9	Q8T6T9 heteractis
31	192	15.0	230	5 Q9GI37	Q9GI37 discosoma s
32	188.5	14.8	227	5 Q95W85	Q95W85 heteractis
33	188	14.7	228	5 Q9GP16	Q9GP16 anemonia su
34	184	14.4	233	5 Q963I9	Q963I9 renilla rec
35	182.5	14.3	238	5 Q9HLZ0	Q9HLZ0 pillosarcus
36	181.5	14.2	231	5 Q9U6Y5	Q9U6Y5 zoanthus sp
37	180.5	14.1	227	5 Q95W86	Q95W86 coelactis
38	174.5	13.7	227	5 Q95W11	Q95W11 coelactis
39	174.5	13.7	231	5 Q8T5E8	Q8T5E8 ricrodea fl
40	172.5	13.5	229	5 Q8T5E7	Q8T5E7 condylactis
41	172	13.5	231	5 Q9U6Y4	Q9U6Y4 zoanthus sp
42	164.5	12.9	231	5 Q8T4U4	Q8T4U4 zoanthus sp
43	162.5	12.7	234	5 Q8C5F3	Q8C5F3 scolyxia cu
44	153.5	12.0	231	5 Q8T6T8	Q8T6T8 discosoma s
45	143	11.2	231	5 Q8T5E9	Q8T5E9 ricrodea fl

ALIGNMENTS

RESULT 1

Q27903 ID Q27903 PRELIMINARY: PRT: 238 AA.
AC Q27903;
CT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9759032; PubMed=9154931;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RC "Enhanced expression in tobacco of the gene encoding green fluorescent
RT protein by modification of its codon usage."
RI Plant Mol. Biol. 33:989-999(1997).
DR EMBL: X96418; CAA65278.1;
DR HSSP: P42212; IGF.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP.1.
DR PRINTS: PR01229; GFLDRESCENT.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26884 MW: CA932D47262AF2D3 CRC64:

Query Match: 98.3%; Score 1254; DB 5; Length 238;
Best Local Similarity 98.3%; Pred. No. 3.2e-97;
Matches 234; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MSKGEELFTGVVPIVLVDG	VNGHKFSVSGEGDATY	GLTKLKFCTTGKILVPPPTL	60
DB	1				
QY	61	VITLSYGVQCFSKRYPDHMKRHDFFK	SAMPEGYVQERTIFFKDDGN	YKTRAEVKEGGDILV	120
DB	61				

```

QY 121 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
DB 122 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238

RESULT 2
Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125; 1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DR 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
AC Aequoreidae; Aequorea.
CX NCBI_TaxID=6100;
RN [1]
RP MEDLINE=96305137; PubMed=8707053;
RX Carmack B.P., Valdivia R.H., Falkow S.;
FA "FACS-optimized mutants of the green fluorescent protein (GFP).";
RT Gene 173:33-38(1996).
RL [2]
RR SEQUENCE FROM N.A.
RA Carmack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.;
RN Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
expression in Candida albicans.";
RL Microbiology 0:0-(1996).
DR EMBL; U73901; AAB18957.1;
DK HSSP; P42212; iBFP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 25940 MW; A2852280A9NEAE CRC64.

Query Match 97.74; Score 1247; DB 5; Length 238;
Best Local Similarity 97.59; Pred. No. 1,2e-95;
Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPVILVELDGVNGHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
DB 1 MSKGEELFTGVVPVILVELDGVNGHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
QY 61 VTTLSYGVOCFSRYPDHMKRHDHFKSAMPGEYGVQERTIFFKDDGNYKTRAEVAFESDITLV 120
DB 61 VTTLSYGVOCFSRYPDHMKRHDHFKSAMPGEYGVQERTIFFKDDGNYKTRAEVAFESDITLV 120
QY 121 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
DB 121 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238

RESULT 3
Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DR 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.

```

```

OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
AC Aequoreidae; Aequorea.
CX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RA Karkios J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X63959; CAA58789.1;
DK HSSP; P42212; iBFP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green_fl_protein; 1.
FT NON-ITER 238
SQ SEQUENCE 238 AA; 26950 MW; 26E2BL450E748E44 CRC64.

Query Match 95.04; Score 1212; DB 5; Length 238;
Best Local Similarity 94.18; Pred. No. 1e-93;
Matches 224; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPVILVELDGVNGHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
DB 1 MSKGEELFTGVVPVILVELDGVNGHKFSVSGEGDGYGKILKFKICTGKLPVFWPTL 60
QY 61 VTTLSYGVOCFSRYPDHMKRHDHFKSAMPGEYGVQERTIFFKDDGNYKTRAEVAFESDITLV 120
DB 61 VTTLSYGVOCFSRYPDHMKRHDHFKSAMPGEYGVQERTIFFKDDGNYKTRAEVAFESDITLV 120
QY 121 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
DB 121 NRLEKGDIFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKVNFKIRHNIEDGGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHVMVLLGFVTAAGITHGMDELYK 238

RESULT 4
Q96J01 PRELIMINARY; PRT; 756 AA.
AC Q96J01;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DR 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Raichu404X.
GN Raichu404X.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN.
RX MEDLINE=21322811; PubMed=11429608;
RA Kocizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
Rapl-1."
RI Nature 411:1065-1068(2001);
DR EMBL; AB051845; BAB01868.1;
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnstrimg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; ras; 1.
DR Pfam; PF02196; RBD; 1.
DR PRODOM; PD013756; Green_fl_protein; 2.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.

```

```

SQ SEQUENCE 758 AA: 85015 MW: 861240826707CFD49 CRC64:
Query Match 94.4% Score 1204: DB 4: Length 758:
Best Local Similarity 95.8% Pred. No. 2,2e-92:
Matches 226: Conservative 4: Mismatches 6: Indels 2: Gaps 0:

QY 1 MSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 60
DQ 499 VSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 558
QY 61 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 120
DQ 559 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 618
QY 121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 180
DQ 619 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 678
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 238
DQ 579 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 734

RESULT 5
Q17106 PRELIMINARY: PRT: 238 AA.
AC Q17106:
QY 01-NOV-1996 (TrEMBLrel. 0): Created)
DT 01-NOV-1996 (TrEMBLrel. 0): Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 2): Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae.
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6109:
RN [1]
RP SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.:
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X63960; CAA58790.1;
DR HSSP: P42212; LBFP.
DR InterPro: IPR000786; Green_fl_protein.
DR PRINTS: PR01229; GFP.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA: 26867 MW: BD4648262D8E8D04 CRC64:

Query Match 93.8% Score 1197: DB 5: Length 238:
Best Local Similarity 93.3% Pred. No. 1.9e-92:
Matches 222: Conservative 7: Mismatches 9: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 60
DQ 1 MSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 60
QY 61 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 120
DQ 61 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 120
QY 121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 180
DQ 121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 238
DQ 181 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 238

RESULT 6
Q8WTC6 PRELIMINARY: PRT: 238 AA.
ID Q8WTC6

```

```

AC Q8WTC6:
DT 01-MAR-2002 (TrEMBLrel. 20: Created)
DT 01-MAR-2002 (TrEMBLrel. 20: Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21: Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae.
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615:
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Gan B.Q., Chen M., Li S.J., Xia N.S.:
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF475431; AAL35916.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27049 MW: 6B5FD75E69326903 CRC64:

Query Match 86.6% Score 1105: DB 5: Length 238:
Best Local Similarity 83.6% Pred. No. 4.5e-85:
Matches 199: Conservative 19: Mismatches 20: Indels 0: Gaps 0:

QY 1 MSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 60
DQ 1 MSKGEELFTGVVPLVLELDGVDNGHKEFSVSGEGSDATYCKLTILKFLICTGKLPVWPPTL 60
QY 61 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 120
DQ 61 VTLISYGVQCFSRYPDMKRRHDFKSAPEGYVQERTIFPKDGNKYKTRAEVKFEGDILV 120
QY 121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 180
DQ 121 NRIELKGIDPKEDGNILGHKLEYNNSHNHYIMADKOKNGIKYVNFIRHNIEDGGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 238
DQ 181 HYQONTPIGDGPVLLPDNHYLSTOSALSQDPNKRDRHNVLLGFVTAAGITHGMDLEYK 238

RESULT 7
Q8WTC6 PRELIMINARY: PRT: 238 AA.
ID Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20: Created)
DT 01-MAR-2002 (TrEMBLrel. 20: Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21: Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae.
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615:
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Gan B.Q., Chen M., Li S.J., Xia N.S.:
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013824; AAK02062.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27049 MW: 8185D0E5E29012B CRC64:

Query Match 86.3% Score 1101: DB 5: Length 238:
Best Local Similarity 83.2% Pred. No. 2e-84:

```

```

Matches 198; Conservative 13; Mismatches 21; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
QY 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
DB 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
DB 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 6
Q8WTC4 PRELIMINARY: PRT: 238 AA.
AC Q8WTC4;
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Cloriflu mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435433; AAL430; A.1.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26956 MW: 75521EAF0CB5A7AA CAG64;
Query Match 96.0%; Score 1097; DB 5; Length 238;
Best Local Similarity 84.0%; Pred. No. 4.4e-84;
Matches 200; Conservative 15; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
QY 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
DB 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
DB 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 9
Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Cloriflu mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435433; AAL430; A.1.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 26956 MW: 75521EAF0CB5A7AA CAG64;
Query Match 85.0%; Score 1055; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNKHKFSVSGEGSDATYGLKLTGKLPVWPPTL 60
QY 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
DB 61 VTTLISYGVQCFSRYPQHKRHDFFKSNPESYVQERTIFFKDDGNKYTRAEVKFEGDILV 120
QY 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
DB 121 NRLEKGDIDFKEDGNLGHKLEYNNSHNHYIMADKKNKNGIKVNEKIRHNIEGGVQLAD 180
QY 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPNHLYSTQSALSKDPNEKRDHMYLLOFVTAAGITHGMDELYK 238
RESULT 10
Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9;
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Cloriflu mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435433; AAL430; A.1.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27031 MW: 5F80A192173CH84D CRC64;
Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;

```

```

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
QY 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
DB 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
QY 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238

RESULT 11
Q8WTC8 PRELIMINARY: PRT: 238 AA.
ID Q8WTC8
AC Q8WTC8
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN-GFPXMI1910V;
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 270-287 MW: 25521EA5534573A CRC64;

Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
RN [1]
RP STRAIN-GFPXMI163;
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 270-287 MW: 255018FA1E7C84D CRC64;

Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
QY 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
DB 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
QY 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238

RESULT 12
Q8WTC7 PRELIMINARY: PRT: 238 AA.
ID Q8WTC7
AC Q8WTC7
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN-GFPXMI163;
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 270-287 MW: 255018FA1E7C84D CRC64;

Query Match 85.7%; Score 1094; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 7.9e-84;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
QY 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
DB 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
QY 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238

```

```

GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN-GFPXMI1910V;
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 270-287 MW: 25521EA5534573A CRC64;

Query Match 85.6%; Score 1092; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 1.2e-83;
Matches 197; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
QY 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
DB 61 VTTLSYGVOGFSRYPDHMKRHDFFKSAHPGQYVQERTIFFKDGDNKYKTRAEVNFPEDTIV 120
QY 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
DB 121 NRIELKGIDFEDKNILGHKLEYNNSHNYVIMADKQKNGIKVNFKIRHNIEDGGVGLAD 180
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
QY 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238
DB 181 HYQONTPTGGPVLPPDHNYLSTQSALSKDPNKRHDHNVLLGFTVTAAGITGHGMDELYK 238

RESULT 13
Q8WTC5 PRELIMINARY: PRT: 238 AA.
ID Q8WTC5
AC Q8WTC5
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN-GFPXMI1910V;
RC Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL: AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 270-287 MW: 75521EA5534573A CRC64;

Query Match 85.4%; Score 1090; DB 5; Length 238;
Best Local Similarity 83.6%; Pred. No. 1.7e-83;
Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDAIVGKLTLSKICTCTTKGKLPVWPPTL 60

```

```

Db      1  MSKGELEFTGVVPLVELDGVGVHKKSVRSGSGADYAKLEIKETCTIKLIPNWPIL 60
QY      61  VTTISYGVQCSRYVDFHMKRHDFKSAPEGYVQERIIFKQDNQNTKTRAEYKPEEDTLV 120
Db      61  VTTISYGVQCSRYVDFHMKRHDFKSAPEGYVQERIIFKQDNQNTKTRAEYKPEEDTLV 120
QY      122  NRLEKGIQDFKEDGNLGHKLEYNNRNHYIMADKQKNGIKYKFXLRINNIHDGGVQLAD 180
Db      121  NRLEKGMDFKEDGNLGHKLEYNNRNHYIMADKQKNGIKYKFXLRINNIHDGGVQLAD 180
QY      181  HYQONTPTCDGPVLLPDHNYLSTOASLKDQNKPRDHMLVLTGTVTAAGLTGCMDELYK 238
Db      181  HYQNTVPLDGPVLPZLNHYLYSTATSKORNETRCHMYLEFPGSGGTHGCMDELYK 238

RESULT 14
Q950A7 PRELIMINARY: PRT; 225 AA.
AC Q650A7
DT 01-DEC-2002 (TREMBLrel. 19, Created)
DI 01-DEC-2002 (TREMBLrel. 19, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cyan fluorescent protein (Fragment).
OS Montastraea cavernosa (Great star coral).
OC Eukaryota; Metazoa; Chordata; Anthozoa; Scleractinia;
OC Pavlovina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
RN [1]
RS SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.J
RA "Montastraea cavernosa fluorescent protein."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ Databases.
DR EMBL; AY056460; AAL17905.1;
DR InterPro; IPR000766; Green_FL_protein.
DR Pfam; PFC1353; GFP.1.
DR ProDom; PD013756; Green_FL_protein.1.
FT NGLTER 225
SC SEQUENCE 225 AA; 52PE2F714D083524 CDS;

```

```

SEQUENCE FROM N.A.
MEDLINE=21927629; PubMed=11929965;
Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F.,
Lukyanov K.A., Lukyanov S.A., Matz M.V.;
" Diversity and evolution of the green fluorescent protein: family.";
Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL: AF420591; AAM10625.1; ...
SQ SEQUENCE 235 AA; 26840 MW; CE1707CFF9334A90 CRC64;

Query Match 19.6%; Score 250.5; DB 5; Length 226;
best Local Similarity 31.8%; Pred. No. 3.2e-13;
Matches 64; Conservative 39; Mismatches 85; Indels 13; Gaps 7

QY 14 ILVE..DGVNKHKSVSGESDATYCKLTKLFCITG-KLPVPWELVTLISYGVGVSFS 72
DB 10 VVXVHEGNNVGHAFVIEEGKGRPYEGTCLNLTIVKGAPLPSYD:ITTAHYGRVET 69
QY 73 RYPDHMKRHDFFKSAHPGYVQERTIEFKDGNYKTRAEVKFEEDTLVNRIELKSTDFKE 132
DB 70 EYPADL--DYFKQSPFGYSEWIMYEDKGICTIRSDLSLGDGFQNIIFNGMNPFP 127
QY 133 DGNILGHK-LEYNYNSHRYIIMADQKNKIV-NFKIRHNIEDGGVQLAHYQONTPIGD 190
DB 128 NGPVMQKTLKPEPSTERLHV----RQGLVGNINMALLGEGHYLADF--KITYAK 180
QY 191 GPVLLPDNNHYLSTQ-SA:SKO 210
DB 181 KVVQIPQYHFVDFVRIEILSD 201

```

RESULT 15	
Q8T600	PRELIMINARY; PRT; 236 AA.
ID	Q8T600
AC	Q8T600;
DT	01-JUN-2002 (TrEMBLrel. 2i, Created)
ET	01-JUN-2002 (TrEMBLrel. 2i, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 2i, Last annotation update)
DE	Green fluorescent protein.
OS	Dendronephthya sp. SSAL-2002
OC	Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonacea;
OC	Nephtheidae; Dendronephthya.
OX	NCBI_TaxID=191210;
RN	{1}

Genome version 5.1.6
Copyright (c) 1993 - 2003 Computer Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:11:38 : Search time 29 Seconds
(without alignments)
974,649 Million cell updates/sec

Title: us-09-967-301-4

Perfect score: 1276
Sequence: 1 MSKGEELFGVPTLVLDG.....VLLRVVANGIHHGMELEK 238

Scoring table:

RLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759772 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:

- 1: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	1262	98.9	238	9	US-10-057-505-2	Sequence 2, Appl
2	1262	98.9	238	10	US-09-884-681-2	Sequence 2, Appl
3	1262	98.9	238	12	US-10-024-636-2	Sequence 2, Appl
4	1262	98.9	1070	9	US-10-001-486-2	Sequence 2, Appl
5	1262	98.9	1099	9	US-10-253-864-6	Sequence 4, Appl
6	1262	98.9	1147	9	US-10-253-864-1	Sequence 1, Appl
7	1262	98.3	1452	12	US-10-050-673-2	Sequence 2, Appl
8	1259	98.7	238	9	US-09-899-954B-2	Sequence 2, Appl
9	1258	98.6	238	9	US-09-899-954B-2	Sequence 2, Appl
10	1258	98.6	238	9	US-09-900-345A-125	Sequence 10, Appl
11	1258	98.6	238	9	US-10-121-258-10	Sequence 10, Appl
12	1258	98.6	238	10	US-10-221-451-6	Sequence 6, Appl
13	1258	98.6	238	10	US-09-920-922-4	Sequence 4, Appl
14	1258	98.6	238	10	US-09-852-900-1	Sequence 4, Appl
15	1256	98.4	244	9	US-09-899-954B-8	Sequence 6, Appl
16	1256	98.4	244	9	US-09-899-954B-10	Sequence 6, Appl
17	1256	98.4	244	9	US-09-899-954B-12	Sequence 12, Appl
18	1256	98.4	244	9	US-09-899-954B-14	Sequence 14, Appl
19	1256	98.4	244	9	US-09-999-954B-16	Sequence 16, Appl

20	1256	98.4	244	9	US-09-999-954B-18	Sequence 18, Appl
21	1256	98.4	238	9	US-09-889-025A-2	Sequence 2, Appl
22	1256	98.4	243	9	US-09-900-345A-60	Sequence 60, Appl
23	1256	98.4	243	9	US-09-900-345A-62	Sequence 62, Appl
24	1256	98.4	243	9	US-09-900-345A-64	Sequence 64, Appl
25	1256	98.4	243	9	US-09-900-345A-66	Sequence 66, Appl
26	1256	98.4	243	9	US-09-900-345A-68	Sequence 68, Appl
27	1256	98.4	243	9	US-09-900-345A-70	Sequence 70, Appl
28	1254	98.3	237	9	US-09-899-954B-36	Sequence 36, Appl
29	1254	98.3	243	9	US-09-900-345A-56	Sequence 56, Appl
30	1254	98.3	243	9	US-09-900-345A-58	Sequence 58, Appl
31	1254	98.3	243	9	US-09-900-345A-114	Sequence 114, Appl
32	1254	98.3	243	9	US-09-900-345A-118	Sequence 118, Appl
33	1254	98.3	243	9	US-09-900-345A-120	Sequence 120, Appl
34	1254	98.3	243	9	US-09-899-954B-4	Sequence 4, Appl
35	1254	98.3	244	9	US-09-899-954B-6	Sequence 6, Appl
36	1254	98.3	244	9	US-09-899-954B-20	Sequence 20, Appl
37	1254	98.3	244	9	US-09-900-345A-2	Sequence 2, Appl
38	1254	98.3	244	9	US-09-900-345A-4	Sequence 4, Appl
39	1254	98.3	244	9	US-09-900-345A-6	Sequence 6, Appl
40	1253	98.2	243	9	US-09-900-345A-8	Sequence 8, Appl
41	1253	98.2	243	9	US-09-900-345A-10	Sequence 10, Appl
42	1253	98.2	243	9	US-09-900-345A-12	Sequence 12, Appl
43	1253	98.2	243	9	US-09-900-345A-14	Sequence 14, Appl
44	1253	98.2	243	9	US-09-900-345A-16	Sequence 16, Appl
45	1253	98.2	243	9	US-09-900-345A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-10-057-505-2
: Sequence 2, Application US/10057505
: Patent No. US20020164674A1
: GENERAL INFORMATION:
: APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
: APPLICANT: AURORA BIOSCIENCES CORPORATION
: APPLICANT: TSIEN, Roger
: APPLICANT: HEIM, Roger
: APPLICANT: CURITT, Andrew
: TITLE OF INVENTION: RANDED FLUORESCENT PROTEIN CONSTRUCTS
: FILE REFERENCE: REG1260-3
: CURRENT APPLICATION NUMBER: US/10/057,505
: PRIOR FILING DATE: 2002-01-25
: PRIOR APPLICATION NUMBER: US 08/792,533
: PRIOR FILING DATE: 1997-01-31
: PRIOR APPLICATION NUMBER: US 09/396,003
: PRIOR FILING DATE: 1999-09-13
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 238
: TYPE: PPT
: ORGANISM: Acqorea victoria
US-10-057-505-2

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 98.7%: Pred. No. 2, 7e-115:

Mismatches 2: Mismatches 1: Mismatches 2: Indexes 0: Gaps 0:

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 98.7%: Pred. No. 2, 7e-115:

Mismatches 2: Mismatches 1: Mismatches 2: Indexes 0: Gaps 0:

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 98.7%: Pred. No. 2, 7e-115:

Mismatches 2: Mismatches 1: Mismatches 2: Indexes 0: Gaps 0:

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 98.7%: Pred. No. 2, 7e-115:

Mismatches 2: Mismatches 1: Mismatches 2: Indexes 0: Gaps 0:

Query Match 98.9% Score 1262: DB 9: Length 238:

Best Local Similarity 98.7%: Pred. No. 2, 7e-115:

Mismatches 2: Mismatches 1: Mismatches 2: Indexes 0: Gaps 0:

QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238
|||||.....
DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238
|||||.....

RESULT 2

US-09-884-681-2

: Sequence 2, Application US/09884681

: Patent No. US20020061546A1

: GENERAL INFORMATION:

: APPLICANT: Tsien, Roger Y.

: CUBIL, Andrew B.

: TITLE OF INVENTION: Assays for Protein Kinases Using

: Fluorescent Protein Substrates

: NUMBER OF SEQUENCES: 48

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Townsend and Townsend and Crew LLP

: STREET: Two Embarcadero Center, Eighth Floor

: CITY: San Francisco

: STATE: California

: COUNTRY: USA

: ZIP: 94111-4834

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Protein Release #1.0, Version #1.3

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/884,681

: FILING DATE: 19-Jun-2001

: CLASSIFICATION: <Unknown>

: Prior APPLICATION DATA:

: APPLICATION NUMBER: 08/675,865

: FILING DATE: <Unknown>

: ATTORNEY/AGENT INFORMATION:

: NAME: Storella, John S.

: REGISTRATION NUMBER: 32,944

: REFERENCE/DOCKET NUMBER: 023072-06-000

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (415) 576-0200

: TELEFAX: (415) 576-0300

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 238 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-884-681-2

Query Match

Best Local Similarity 98.9%; Score 1262; DB 10; Length 238;

Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTLVLDGVNGHKFVSQSGEGDGYGKLTFRFICTIGKLPVWPPL 60

DB 1 MSKGEELFTGVPTLVLDGVNGHKFVSQSGEGDGYGKLTFRFICTIGKLPVWPPL 60

QY 61 VTILSYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEEDTLV 120

DB 61 VTILSYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEEDTLV 120

QY 121 NRILKIDFKEDNTLGHKLEYNYNHNYIMADKQNGIKVNFKIRNIEGSGVOLAD 180

DB 121 NRILKIDFKEDNTLGHKLEYNYNHNYIMADKQNGIKVNFKIRNIEGSGVOLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238

RESULT 3

US-10-024-686-2

: Sequence 2, Application US/10001495B

: Publication No. US20030082642A;

: Sequence 2, Application US/10024686

: Patent No. US20022123113A;

: GENERAL INFORMATION:

: APPLICANT: Tsien, Roger Y.

: Helm, Roger

: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

: NUMBER OF SEQUENCES: 5

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Fish & Richardson P.C.

: STREET: 4225 Executive Square, Suite 1400

: CITY: La Jolla

: STATE: CA

: COUNTRY: USA

: ZIP: 92037

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: Windows95

: SOFTWARE: FastSeq for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/10/024,686

: FILING DATE: 17-Dec-2001

: Prior APPLICATION DATA:

: APPLICATION NUMBER: 09/557,955

: FILING DATE: <Unknown>

: APPLICATION NUMBER: 06/727,452

: FILING DATE: 10-OCT-1996

: APPLICATION NUMBER: US95/14652

: FILING DATE: 13-NOV-1995

: APPLICATION NUMBER: 08/337,915

: FILING DATE: 10-NOV-1994

: ATTORNEY/AGENT INFORMATION:

: NAME: Haile, Lisa A.

: REGISTRATION NUMBER: 35,347

: REFERENCE/DOCKET NUMBER: 07257/032022

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 619/678-5070

: TELEFAX: 619/678-5099

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 238 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

: FRAGMENT TYPE: internal

: SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-024-686-2

Query Match

Best Local Similarity 98.9%; Score 1262; DB 12; Length 238;

Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPTLVLDGVNGHKFVSQSGEGDGYGKLTFRFICTIGKLPVWPPL 60

DB 1 MSKGEELFTGVPTLVLDGVNGHKFVSQSGEGDGYGKLTFRFICTIGKLPVWPPL 60

QY 61 VTILSYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEEDTLV 120

DB 61 VTILSYGVQCFSRYPDHMKRDEDFKSAMPEGYVQERTIFFKDDGNYKTRAEYKFEEDTLV 120

QY 121 NRILKIDFKEDNTLGHKLEYNYNHNYIMADKQNGIKVNFKIRNIEGSGVOLAD 180

DB 121 NRILKIDFKEDNTLGHKLEYNYNHNYIMADKQNGIKVNFKIRNIEGSGVOLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALSADPNKPKRHHVLLFEVTAAGITHGMDELYK 238


```

GENERAL INFORMATION:
APPLICANT: The Government of the United States of America
as represented by the Secretary
Department of Health and Human Services
Washington, D.C.
Htun Ph.D., Han
Hager Ph.D., Gordon L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MANIPULATING
DNA BINDING MOLECULES IN LIVING CELLS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.00
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/16/001,486B
FILING DATE: 15-Nov-2003/052642A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,373
FILING DATE: 08-Dec-1995
ATTORNEY/AGENT INFORMATION:
NAME: Selby, Elizabeth
REGISTRATION NUMBER: 38298
REFERENCE/DOCKET NUMBER: 14014.0183
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-688-0770
TELEFAX: 404-688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-601-488-2
Query Match 98.9% Score 1262; DB 9; Length 1079;
Best Local Similarity 98.7% Pred. No. 2,1e-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
DB 1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
QY 61 VTITTYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61 VTITTYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
QY 121 NRTELKGIQKEDGNILGHKLEYNNSHNVYIMAKQKNGIKVNFIRHNLEDGVSQVLAD 160
DB 121 NRTELKGIQKEDGNILGHKLEYNNSHNVYIMAKQKNGIKVNFIRHNLEDGVSQVLAD 160
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVVLEFVTAAGITHGMDFLYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDHVVLEFVTAAGITHGMDFLYK 238
RESULT 5
US-10-259-864-1
Sequence 1, Application US/10/259864
Publication No. US20030077645A1
GENERAL INFORMATION:
APPLICANT: Hager, Gordon L.
TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfa
Receptor Ligands, and Methods and Kits For Detecting and Char
TITLE OF INVENTION: Ligands
FILE REFERENCE: 24799
CURRENT APPLICATION NUMBER: US/10/259,864
PRIOR FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US 60/325,178
PRIOR FILING DATE: 2001-09-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 1
LENGTH: 1147
TYPE: PPT
ORGANISM: Rat/human chimera
FEATURE:
NAME/KEY: mat_peptide
LOCATION: (1)..(1)
OTHER INFORMATION: Chimeric protein
US-10-259-864-1
Query Match 98.9% Score 1262; DB 9; Length 1147;
Best Local Similarity 98.7% Pred. No. 2,2e-114;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
DB 1 MSKGEELFTGVVPLVELDGDVNGHKFSVSGEGDATYGKLTLEFICTGKLPVWPTL 60
QY 61 VTITTYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61 VTITTYGVQCFSRYPDHMKRHDFFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120

```


Sequence 125, Application US/05950345A
 Publication No. US20030031999A1
 GENERAL INFORMATION:
 APPLICANT: Frazer, Ian Hector
 APPLICANT: Zhou, Jian
 TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
 EFFICIENCY OF A CODON
 FILE REFERENCE: 10338-5US
 CURRENT APPLICATION NUMBER: US/05/960,345A
 PRIOR FILING DATE: 2001-07-06
 PRIOR APPLICATION NUMBER: AU PP4278
 PRIOR FILING DATE: 1999-01-06
 PRIOR APPLICATION NUMBER: PCT/AU00/00008
 PRIOR FILING DATE: 2000-01-07
 NUMBER OF SEQ ID NOS: 185
 SOFTWARE: Patents: Ver. 2.0
 SEQ ID NO 125
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: GFP Humanized
 OTHER INFORMATION: control
 US-09-967-301-4.rapb

Query Match 98.6%; Score 1258; DB 9; Length 238;
 Best Local Similarity 98.3%; Pred. No. 6.6e-115;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 DB 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 QY 61 VTLLTYGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 DB 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238
 DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238

Query Match 98.6%; Score 1258; DB 9; Length 238;
 Best Local Similarity 98.3%; Pred. No. 6.6e-115;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 DB 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 QY 61 VTLLTYGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 DB 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238
 DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238

QY 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 DB 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 QY 61 VTLLTYGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 DB 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238
 DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238

RESULT 12
 US-10-221-461-6
 Sequence 6, Application US/10221451
 Publication No. US20030092902A1
 GENERAL INFORMATION:
 APPLICANT: Marsh, Donald J.
 TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR
 FILE REFERENCE: 20652P
 CURRENT APPLICATION NUMBER: US/10/221,461
 CURRENT FILING DATE: 2002-09-12
 PRIOR APPLICATION NUMBER: PCT/US01/08071
 PRIOR FILING DATE: 2001-03-14
 PRIOR APPLICATION NUMBER: 60/189,598
 PRIOR FILING DATE: 2000-03-15
 NUMBER OF SEQ ID NOS: 37
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea Victoria
 US-10-221-461-6

Query Match 98.6%; Score 1258; DB 9; Length 238;
 Best Local Similarity 98.3%; Pred. No. 6.6e-115;
 Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 DB 1 MSKGELFTGVVPLVLELDGDNVNGHKFSVSGEGSDATYKLTAKFICITTKGLPVPWPTL 60
 QY 61 VTLLTYGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTFSGVOCFSRYPDHKKRHHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 DB 121 NRLEKSIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVLAD 180
 QY 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238
 DB 181 HYQONTPIGCGPVLLPDNHYLSTQSALSCKDPNEKRDHVVLEFVTAAGITRGMDELYK 238

RESULT 13
 US-09-920-922-4
 Sequence 4, Application US/9920922
 Patent No. US2002008448A1
 GENERAL INFORMATION:
 APPLICANT: Miyawaki, Atsushi
 APPLICANT: Sawano, Asako
 TITLE OF INVENTION: METHOD FOR MUTAGENESIS
 FILE REFERENCE: 11283-012001
 CURRENT APPLICATION NUMBER: US/09/920,922
 CURRENT FILING DATE: 2001-08-02

? PRIOR APPLICATION NUMBER: JP 2000-237166
 ? PRIOR FILING DATE: 2000-08-04
 ? NUMBER OF SEQ ID NOS: 9
 ? SOFTWARE: FastSeq for Windows Version 4.0
 ? SEQ ID NO 4
 ? LENGTH: 238
 ? TYPE: PRT
 ? ORGANISM: Aequorea victoria
 ? US-09-920-922-4

Query Match 98.6%; Score 1258; DP 10; Length 238;
 Best Local Similarity 98.3%; Pred. No. 6.6e-115;
 Matches 234; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 VTTLTYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VTTSYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

RESULT 14

US-09-952-000-1

Sequence 1, Application US/09852400

Patent No. US20020099170A1

GENERAL INFORMATION:

APPLICANT: Osumi, Takashi

APPLICANT: Tsukamoto, Toshio

APPLICANT: Tsukamoto, No. US20020099170A1

APPLICANT: Yamasaki, Masahiko

TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT

TITLE OF INVENTION: PROTEINS

FILE REFERENCE: 040124-5005-US

CURRENT APPLICATION NUMBER: US/09/852,000

PRIOR FILING DATE: 2001-03-10

PRIOR APPLICATION NUMBER: JP 024418/1998

PRIOR FILING DATE: 1998-01-23

PRIOR APPLICATION NUMBER: US 09/121,539

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 09/515,655

PRIOR FILING DATE: 2000-07-13

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 1

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria

FEATURE:

OTHER INFORMATION: Green fluorescent protein

US-09-852-000-1

Query Match 98.6%; Score 1258; DP 10; Length 238;
 Best Local Similarity 98.3%; Pred. No. 6.6e-115;
 Matches 234; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 VTTLTYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VTTSYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

```

QY 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-899-954B-8
Sequence 8, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
TITLE OF INVENTION: Target cell or tissue of a plant
FILE REFERENCE: 10338-6US
CURRENT APPLICATION NUMBER: US/09/999,954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PP8077
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCT/AU00/0007
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patent in version 3.1
SEQ ID NO 6
LENGTH: 244
TYPE: PRT
FEATURE:
ORGANISM: Artificial Sequence
OTHER INFORMATION: Lm(CTA)6 GFP construct
US-09-899-954B-8
  
```

Query Match 98.4%; Score 1256; DP 9; Length 244;
 Best Local Similarity 97.9%; Pred. No. 1.1e-114;
 Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MSKGEELFTGVVPIVLVDGNGVNGHKFSVSGEGEDATYGLTKLTKFCTTCKLPPVWPTL 66
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 VTTLTYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VTTSYGVQCFSRYPDHRKHDFKSAPEGVVQERTIFFKDDGNKYKTRAFVKFEGDTLV 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 NRLELKGIDFKEDGNILGHKLEYNNSHNHYIMADKQKNGIKVNFKIRHNIEDGGVOLAD 186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 HYQQNTPIGDGPVLLPDNHYLSTOSALSADPNKRDHNVLEFVTAAGTTHGMDELK 244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

Search completed: July 16, 2003, 16:30:11

Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer M.I.

CM protein - protein search, using sw model:

Run on: July 16, 2003, 16:07:32 : Search time 25.6667 seconds
(without alignments)
1:07.296 Million cell updates/sec

Title: us-09-967-301-4

Perfect score: 1216

Sequence: 1 MKSGEETGVVPIIVLELGG.....VLIEPVAAATIMHMDYK 23P

Scoring table: BKSUM62

Gapop 10.0 : Gapext 0.5

Searched: 281224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 281224

Minimum DB seq length: 0
Maximum DB seq length: 2000003000

Post-processing: Minimum Match 04
Maximum Match 1008
Listing first 45 summaries

Database : PIR_73:.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252	98.1	238	1 J01514	green-fluorescent
2	100	7.8	785	2 H72228	hypothetical prote
3	99.5	7.7	887	2 H82590	leucyl-tRNA synth
4	92.5	7.2	534	1 N1CJMA	nitrogenase (EC 1.
5	92	7.2	336	2 C64468	hypothetical prote
6	91	7.1	439	2 JH0414	synaptobactin G-65
7	89	7.0	1224	1 E9H0A8	cholesterol complex
8	88.5	6.9	861	2 H64102	leucine-tRNA ligas
9	88	6.9	531	2 C95338	hypothetical prote
10	87.5	6.9	370	2 E70390	iron-sulfur cent
11	87	6.6	357	2 G81355	tRNA (Glu-tl)-5-m
12	86.5	6.8	222	1 A15028	DNA-directed DNA p
13	86.5	6.8	2573	2 D71614	hypothetical prote
14	86	6.7	481	2 AD2052	hypothetical prote
15	85.5	6.7	263	2 S53458	water-stress-induc
16	85.5	6.7	700	1 C1HCH2	galbanin (EC 3.4.22
17	85.5	6.7	797	2 F01079	protective surface
18	85.5	6.7	806	2 F64102	protective surface
19	85.5	6.7	860	2 A05882	leucyl-tRNA synth
20	85	6.7	47	2 T27556	hypothetical prote
21	85	6.7	632	2 T05586	DNA-binding protei
22	84.5	6.6	564	2 F81177	ABC-type transpor
23	84.5	6.6	655	2 D83917	DNA replication
24	84.5	6.6	884	2 J05576	inter-alpha-trypsi
25	83.5	6.5	422	2 B24813	calpain (EC 3.4.22
26	82.5	6.5	312	2 C81713	hydroxylase reduct
27	82.5	6.5	1134	2 A60234	EGF receptor pr
28	82.5	6.5	1164	1 F05043	EGF receptor pr
29	82	6.4	470	2 T31349	hypothetical prote

30	81.5	6.4	330	2 C90208	hypothetical prote
31	81.5	6.4	513	2 A99552	oligodehydrogenase
32	81.5	6.4	536	2 C70031	hypothetical prote
33	81.5	6.4	788	1 GDLVHH	DNA-directed DNA p
34	81.5	6.4	2518	2 A12140	polyketide synthas
35	81	6.3	422	1 BMS72Y	synaptobactin II -
36	81	6.3	547	2 H89988	hypothetical prote
37	81	6.3	574	2 J04930	S-layer protein pr
38	80.5	6.3	389	2 G84245	NADH dehydrogenase
39	80.5	6.3	393	2 C64613	conserved hypothet
40	80.5	6.3	752	1 KXPTC1	proprotein convert
41	80.5	6.3	1259	2 AE1055	probable exported
42	80	6.3	353	2 F84941	imidazoleglycerol d
43	80	6.3	578	1 I40794	probable oxysterol
44	80	6.3	760	2 E84715	hypothetical prote
45	80	6.3	764	2 T25012	hypothetical prote

ALIGNMENTS

RESULT 1

J01514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence,revision 03-Dec-1999 #extl_change 23-Mar-2001
C:Accession: J01514; PQ0335; S48693; S51330; S51331
R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cozart, M.C.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: J01514; M01D:92175527; PMID:1347277
A:Accession: J01514
A:Molecule type: DNA
A:Residues: 1-307, S, 109-236 <PRA1>
A:Cross-references: GB:M62654; NID:q155662; P1DN:AAA27722.1; P1D:q155663
A:Accession: J01514
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GB:M62653; NID:q155660; P1DN:AAA27722.1; P1D:q155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64, 74-122, 132-151, 154-183, 185-200 <PRA3>
R:Imoye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A:Reference number: S48693; M01D:94364470; PMID:8082767
A:Accession: S48693
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>
A:Cross-references: GB:L29345; NID:q606383; P1DN:AAA58246.1; P1D:q606384
R:Watkins, J.N.; Campbell, A.K.
Submitted to the EMBL Data Library, January 1995
A:Reference number: S51330
A:Accession: S51330
A:Molecule type: mRNA
A:Residues: 1-13, 'V', 15-24, 'Q', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227, 'A:Cross-references: EMBL:X83900; NID:q634008; P1DN:CAA58799.1; P1D:q634009
A:Experimental source: clone gfp
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A65892; P081GFI.
A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A', 2-93, 'E', 8
A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli
R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.
Nat. Biotechnol. 14, 1246-1251, 1996
A:Title: The molecular structure of green fluorescent protein.
A:Reference number: A58953; M01D:98294543; PMID:9631087

A:Contents: annotation: X-ray crystallography, 1.3 angstroms
 C:Comment: This protein is excited by the photoreactive region (see P14607FV) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-66/hydro-Tyr
 C:Genetics:

A:Gene: GFP
 A:Introns: 69/3; 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/cross-link: 5-imidazolilone (Ser-Gly) #status experimental
 F:66/modified site: dehydrotyrosine (Tyr) #status experimental
 Query Match 98.1%; Score 1252; DB 1; Length 238;
 Best Local Similarity 97.1%; Pred No. 2.4e-93;
 Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELSDVNGKFSVSGEGGDAATYKLUKFKCTCTCKLVPVPTL 60
 DB 1 MSKGEELFGVVPILVELSDVNGKFSVSGEGGDAATYKLUKFKCTCTCKLVPVPTL 60
 QY 61 VTITLVGVCFGRYPDHMKRHDFFKSAKPEGVVQERTIFPKDGNKTPAEVKPEGDI 120
 DB 61 VTITLVGVCFGRYPDHMKRHDFFKSAKPEGVVQERTIFPKDGNKTPAEVKPEGDI 120
 QY 121 NRIELKGDIFKEDGNILGHKLEYNNSHNYVINALKQKNTKYNKKEIRNTEDSGVOLAD 180
 DB 121 NRIELKGDIFKEDGNILGHKLEYNNSHNYVINALKQKNTKYNKKEIRNTEDSGVOLAD 180
 QY 181 HYQNTPIGDDGVLLPDNIYLSQSLSKDPEKQKQKQKQKQKQKQKQKQKQKQKQK 236
 DB 181 HYQNTPIGDDGVLLPDNIYLSQSLSKDPEKQKQKQKQKQKQKQKQKQKQKQKQ 236

RESULT 2
 R72228
 hypobothical protein TM1624 - Thermotoga maritima (strain MS84)
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2003
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.P.; Oviatt, M.; Dodson, R.; Haft, D.B.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, K.S.; Phillips, G.A.; Richardson, P.
 C.M.

Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72228; MID:99287416; PMID:10469571
 A:Accession: H72228
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-785 <AP>
 A:Cross-references: GB:AE000502; GB:AE000512; NCBI:gb|F199:110449+219
 A:Experimental source: strain MS86
 C:Genetics:
 A:Gene: TM1624

Query Match 7.8%; Score 100; DB 2; Length 785;
 Best Local Similarity 20.4%; Pred No. 1.8;
 Matches 40; Conservative 27; Mismatches 54; Indels 6; Gaps 0;
 QY 2 SKGELFTGVVFLVSDGVNGKFSVSGEGGDAATYKLUKFKCTCTCKLVPVPTL 41
 DB 15 NEGKFSFGVPGVQWQAD 46
 QY 62 TLTLYGVQCFGRYPDHMKRHDFFKSAKPEGVVQERTIFPKDGNKTPAEVKPEGDI 121
 DB 47 -----NEDLFKEDREWIYEREFKEDKESKERVVDVFEVGVGLS 88
 QY 122 RIELKGDIFKEDGNILGHKLEYNNSHNYVINALKQKNTKYNKKEIRNTEDSGVOLAD 181
 DB 89 DVTYLVGVLL-----GSTEDMFLYEFVDINVL-----KEKNILKXVYK-----SPKRVKPT 134
 QY 182 YQNTPIGDDGP 192
 DB 135 LEQNYGVVGGP 145

RESULT 3
 E82590
 leucyl-tRNA synthetase XF2176 (imported) - Xylella fastidiosa (strain 945c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: E82590
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 405, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59128 below
 A:Accession: E82590
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-887 <Sim>
 A:Cross-references: GB:AE004031; GB:AE003849; NCBI:gb|0107309:1; GSPB:GN
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accacio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Cartao, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Doroty, H.; Facincani, A.P.; Ferreira, A.C.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 C.D.; Junqueira, M.L.; Komper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.F.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., K.A.; da Silv
 M.; Tschako, M.H.; Vahlada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.C.

A:Reference number: A59128
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2176
 C:Superfamily: leucine-tRNA ligase

Query Match 7.7%; Score 98.5; DB 2; Length 887;
 Best Local Similarity 23.2%; Pred No. 2.8;
 Matches 46; Conservative 27; Mismatches 68; Indels 57; Gaps 10;
 QY 49 ITGKLPVPTLVITLVGVCFGRYPDHMKRHDFFKSAKPEGVVQERTIFPKDGNV 106
 DB 329 INGLPV-WYANVIVMAYVIGAVMAYPGHQRQFF--ANKYGLPIQVIA:KEPKNLE 385
 QY 107 -----KTRAEVKPEGDIIVNRIRIKIGDIFKEDGNILGHKLEYNNSHNYV 152
 DB 146 STWEDPVWQWYADKTR--EEF---LINSAEEDSDYQDAFEVLAERFE--- 429
 QY 153 MADKQKNG-IKYNEKIRNNTEDSGVOLADHYQNTPI-----SIGKPVLPEN 196
 DB 430 ---RQGGQKQWYRLP--DWGNSRQRYWQCPVITVTCGAVPVFELQIPVLPEN 482
 QY 199 -HYLSTQSALSQKDFEKA 215
 DB 483 VAFSGTGSPIKTDPEK 500

RESULT 4
 NICLMA
 Nitrogenase (EC 1.18.6.1) molybdenum-iron: protein alpha chain [validated] - Clostridi
 N:Alternate names: dinitrogenase alpha chain; nitrogenase component I alpha chain
 C:Species: Clostridium pasteurianum
 C:Date: 01-Sep-1981 #sequence_revision 31-Dec-1991 #text_change 19-Jan-2001
 C:Accession: S07389; I40816; A00545
 R:Wang, S.Z.; Chen, J.S.; Johnson, J.L.
 Nucleic Acids Res. 15, 3935, 1987
 A:Title: Nucleotide and deduced amino acid sequences of nitC encoding the alpha-subu
 A:Reference number: S07389; MID:97231095; PMID:3473447
 A:Accession: S07389
 A:Molecule type: DNA
 A:Residues: 1-534 <WAN>
 A:Cross-references: EMBL:Y00155; NCBI:gb|0583:1; PID:q580995
 R:Chen, K.C.


```
A:Gene: Sma1131
A:Genome: plasmid

Query Match      6.8%  Score 88:  DB 2:  Length 532:
Best Local Similarity 20.4%:  Pred. No. 11:
Matches 61:  Conservative 29:  Mismatches 27:  Indels 122:  Gaps 14:

QY 6 ELFTGVVPLVLEGGVNV-CHKFSVSGEGEDATYKGLDKFKICIGKLVFNPVLTATL 64
DB 176 EITGSPKRLFLSGDLSRPHKL-----GGP-  BAKAFSPKQDVICHS 216
QY 65 TYGVQCFSPYPMKKHH-----DFKFSAMFEG-----YVLEPK-----IFPK----- 162
DB 217 TYGTEREADGAARPHILRSEVLTAHPNGALLSPSPAFVFTQRLIEIVHMEGSAVP 276
QY 203 -----GGNFKTRAEVKF-----EGDTLVNKEILKSLRKE----- 182
DB 277 KCFILDSPLATRASIFRHHARELNGALVATSKNVPTTAFQSRVALLRHHF 336
QY 233 -----DGNLGHKLE-YNYNSHNVYIMADCKND----- 149V 164
DB 337 VIANSGMCEAGRIERIKKWLARDEATVLLVGYDANCLLELLEATVAVVAGDMPYR 466
QY 165 FKPH-NIEG---GVLADHIE-NTFIDGSPV----- 162
DB 397 AR-EKLDIVSHADGSLADKVVARGQVRAGVNVHGHGNEKIFERLSTEFQVHED 454

RESULT 10
E70390
Iron-sulfur cofactor synthesis protein nifs Aquifex aeolicus
N:Contents: L-cysteine sulfurtransferase (EC 2.8.1.1)
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 07-Dec-1999
C:Accession: E70390
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lerox, A.L.; Graham, L.E.; O'
V.
Nature 392, 353-358, 1999
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; PMID:98195666; PMID:9537329
A:Accession: E70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1,370 <AO>
A:Cross-references: GB:AE000720; NID:q2983529; P1DN:AA037111; P1D:q2954536; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: nifs1
A:Superfamily: nitrogen fixation protein nifs
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
F:195/Binding site: pyridoxal phosphate (lys) (covalent) *status predicted
F:319/Active site: cys (cysteine persulfide intermediate) *status predicted

Query Match      6.8%  Score 67.5:  DB 2:  Length 370:
Best Local Similarity 25.4%:  Pred. No. 7:
Matches 49:  Conservative 30:  Mismatches 9:  Indels 23:  Gaps 6:

QY 3 KGEELFTGVV-----PIHVELD---GVNKHKF-SVSEH---EGVATYGLKLEKFI 67
DB 164 KGVPLTDAVQAQCKPIE-LKNSYATFSHKHFKAGSPHATSEANVEPLIVGGGG 224
QY 50 TGKLP-----VPMPTLVTTITVGVQFSRYPMHKK-HHFKFSAMFERYVQERTIFPK 109
DB 224 NGRKSGTERVVGILSLAKALELIVSNFSRYOELKKLRLDFLNFLEA-LPDAQVKKGA 282
QY 104 GNVKTRAEV---AFKQDTLVNRLKSLDKFEGNGLGHKLEYNYNSHNVYIMADCKND 160
DB 283 ERSPSSISVMPAPFGAGIYNKLSKSGVCS-LSVALSDIYFENKMKLNKMFQVKALAK 342
QY 161 IKVNFKRRHIED 173
DB 343 VRPSFGLLNKEE 355

RESULT 11
G81355
tRNA (cylcyl-5)-methyltransferase (EC 2.1.1.35) Cj0831c [imported] - Campylobacter j
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jul-2002
C:Accession: G81355
R:Parkhill, J.; Wren, B.W.; McDall, K.; Kottley, J.M.; Churcher, C.; Basham, D.; Chl
C.W.; Ouail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bai
Nature 403, 665-669, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: G81355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <PA>
A:Cross-references: GB:AL111159; NID:q966128; P1DN:CAH73096.1; P1D:q9696
A:Experimental source: serotype O2, strain NCIC 11108
C:Genetics:
A:Gene: trmA; Cj0831c
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match      6.8%  Score 87:  DB 2:  Length 357:
Best Local Similarity 24.8%:  Pred. No. 8:
Matches 30:  Conservative 16:  Mismatches 39:  Indels 54:  Gaps 9:

QY 79 KRHDFKSAPEGYVQENTIFSKDIGNYKTRAEVKF---EGDTLV-----NPIELKG 127
DB 14 EKRSFKKFKFEFTYKQKLPASKDKHYKRAELSFYHENTILFYAMDHKSKKYLLEY 73
QY 128 DFKED-----GNLGHKLEYNYNSHNVYIMADCKNGIKYKFKIRHNE 172
DB 74 LDFADEKICAEFRILEYLRQDNK-LKEKI-----EGVEFTIKCE--LSITLIVHKNLE 125
QY 173 D 173
DB 126 D 126

RESULT 12
A36028
DNA-directed DNA polymerase (EC 2.7.7.7) II catalytic chain - yeast (Saccharomyces ce
N:Alternate names: DNA-directed DNA polymerase II chain A; protein N0825; protein YNL
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A36028; B36028; S60919; S63235; S65121
R:Morrisson, A.; Araki, H.; Clark, A.B.; Hamatake, R.K.; Sugino, A.
Cell 62, 1143-1151, 1990
A:Title: A third essential DNA polymerase in Saccharomyces cerevisiae.
A:Reference number: A36028; MUID:90381771; PMID:2169349
A:Accession: A36028
A:Molecule type: DNA
A:Residues: 1-2222 <MO>
A:Cross-references: GB:M60416; GB:M36724; NID:q171408; P1DN:AAH8711.1; P1D:q171409
A:Accession: B36028
A:Molecule type: protein
A:Residues: 1214-1216 'X', 1218-1221 <MO>
R:Sen-Gupta, M.; Lyck, P.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.H.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV
A:Reference number: S60909
A:Accession: S60919
A:Molecule type: DNA
A:Residues: 1-2221 <SEN>
A:Cross-references: EMBL:X52494; NID:q1045236; P1DN:CAH63235.1; P1D:q1045247
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence database, April 1996
A:Reference number: S63235
A:Accession: S63235
A:Molecule type: DNA
A:Residues: 1-2222 <SEN>
A:Cross-references: EMBL:Z71538; NID:q1302316; P1DN:CAA646169.1; P1D:q1302317; GSPDH:G
A:Experimental source: strain S288C
```

R/Son-Gupta, M.; Byck, R.; Field, G.; Niedenthal, R.K.; Regenstein, J.H.

Yeast 22, 505-514, 1996

A:Title: The sequence of a 24 kb segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae* strain 1024-11C is identical to the sequence of the left arm of chromosome XIV from *S. cerevisiae* strain 1024-11C

A:Reference number: S65111; MID:96310631; PMID:8740425

C:Accession: S65121

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2221 <SEF>

A:Cross-references: EMBL:X82434; NID:91045236; PIR:CA664245.1; PMID:81045247

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Gene: SGD:POL2; DUN2; MIPS:YNL262w

A:Cross-references: SGD:S0005206; MIPS:YNL262w

A:Map position: 14L

C:Superfamily: DNA-directed DNA polymerase II

C:Keywords: DNA binding; nucleotidyltransferase; nucleus; zinc finger

Query Match 6.8% Score 86.5; DB 2; Length 2222;

Best Local Similarity 28.2%; Pred. No. 97;

Matches 37; Conservative 14; Mismatches 45; Indels 31; Gaps 7;

QY 53 LPVWP-ILVTTLLTYGVQCSRPEDH-----KREIFFKSAMPGV-----SEPTI 96

DB 883 LKSFDEIYFLLNGKKLYSPCSMLNYRVCKFTNRYGELADFLNLYTHSENTI 942

QY 99 FFKDDGNYKTR--AEVKFEGDTLVNR-----IEKSGIDQKFGNLTGHKLEYN 146

DB 943 EFEVDGPKAMILPSSKEEGKGIKKYAVNHLGSLAEKSPKPKAGE--QILKNVQ 999

QY 147 S--HNVTIMAD 155

DB 1000 SDIPKVPLEGD 1010

RESULT 13

D71614

Hypothetical protein PF00460c - malaria parasite (*Plasmodium falciparum*)

C:Species: *Plasmodium falciparum*

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: D71614

R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.K.; Arayind, L.; Koonin, E.V.;

Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.B.; Clayton, R.; White, G.; Smith, R.B.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.

A:Reference number: A71600; MID:95221743; PMID:9604761

A:Accession: D71614

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2573 <QAR>

A:Cross-references: GB:AE001396; GS:AE001392; NID:93445168; PIR:AA71382.1; EMBL:438471

A:Experimental source: clone 307

C:Genetics:

A:Gene: PF00460c

Query Match 6.8% Score 86.5; DB 2; Length 2573;

Best Local Similarity 26.2%; Pred. No. 112c-02;

Matches 34; Conservative 29; Mismatches 54; Indels 14; Gaps 5;

QY 93 VQERTTFFKQ--SGNYKTRAEVFEFGDTLVNRIELKGLDGFKEGNIHGLKLEY--YKSH 148

DB 126 LKRSITLLCKD:KSGSDPMDETSLSKIDMYDQRECK--DEKSKSKKREVFNTYNNH 204

QY 149 NVYIMADKQKNGKVNFKIRENTEDGSGVLADHYVOONTPTGQVLPONHLSLSALS 208

DB 184 NLRIKENKDKQKFKKKN:HNNDNNNNK---IYKMI---DKTHILDNNVVELLGIN 246

QY 209 KDPNKKRDMH 218

DB 237 TYLKRBDYM 246

RESULT 14

AD2052

Hypothetical protein all1970 (imported, - Nostoc sp. (strain PCC 7120)

C:Species: *Nostoc sp.*

A:Note: *Nostoc sp.* strain PCC 7120 is a synonym of *Anabaena sp.* strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AD2052

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, I.; Sasamoto, S.; Katahata, A.; Iriki

Nakazaki, N.; Shirao, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium

A:Reference number: AB1807; MID:21595285; PMID:11759840

A:Accession: AD2052

A:Molecule type: DNA

A:Residues: 1-261 <KOR>

A:Cross-references: GB:BA000019; PIR:BA073659.1; PIR:51711040; SPS05:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all1970

Query Match 6.7% Score 85.5; DB 2; Length 263;

Best Local Similarity 22.3%; Pred. No. 7.2;

Matches 46; Conservative 21; Mismatches 86; Indels 51; Gaps 6;

QY 31 GGGEDATYVGLTLAFICTGKLPVWPPTLVITLVITGVQCFSRFLDKMKRHFFKSAKPE 90

DB 41 GSEKSTSYGKTSYSGDTYGE-----KTTFGDD--NKYCKTSYGLTYGKPT 79

QY 91 GY-----VQERTTFFKQDQNYKTRAEVFEFGDTLVNRIELKGLDGFKEGNIHGLKLEY-- 143

DB 80 SYGGDNTYGEKTSYSGKDKNYGKTSYSGE-----DGNKYGKTSYSG 123

QY 144 -NYSNHYIMADKQKNGIKVNFKIRHIEDGGVLADHYVOONTPTGSGVLLPNNHLS 202

DB 124 SYGKPSYGGSDKNYGEKTSYSGNEEGYGGVGGETTNYEN-----E 167

QY 203 TGSALSKDPNKKRGMVLEFVTAAG 228

DB 169 SETKTSCKYKBEKKHKKHLEIGGLG 193

Thu Jul 17 09:57:18 2003

us-09-967-301-4.rpr

Page 7

Search completed: July 16, 2003, 16:14:40
Job time : 21.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search using sw mode

Run on: July 16, 2003, 16:11:23 : Search time 15.6667 seconds
(with 3.3 alignment/s)
926,441 Matches (all) of 4409/500

Title: US-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKGEELPTGVVPLVLELQ3.....VILLEVTAAGLTHQHELYE 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searches: 112692 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 12562

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 04
Maximum Match 1004
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1258	98.6	238	1	GFP_ARQV2
2	98.5	7.7	879	1	SYL_XYLFA
3	92.5	7.2	533	1	NIFD_CLOPA
4	92	7.2	336	1	YD48_MEIJA
5	91	7.1	439	1	SV62_D_SOM
6	89	7.0	1224	1	COPA_HUMAN
7	88.5	6.9	861	1	SYL_HAFIN
8	87	6.8	357	1	TRMA_CAMEL
9	86.5	6.8	2222	1	DPUE_YEAST
10	86	6.7	1224	1	COPA_BOVIN
11	85.5	6.7	700	1	CAN2_HUMAN
12	85.5	6.7	795	1	D152_HAFIN
13	85.5	6.7	797	1	D151_HAFIN
14	84.5	6.6	324	1	CAN2_PIG
15	84.5	6.6	793	1	D153_HAFIN
16	84.5	5.6	886	1	ITH_MESARJ
17	83.5	6.5	422	1	CAN2_RABIT
18	83.5	6.5	533	1	CP51_CANGA
19	82.5	6.5	312	1	TRXS_CHLMC
20	82.5	6.5	1164	1	ARG_STKAG
21	81.5	6.4	890	1	PRBL_SULSO
22	81.5	6.4	613	1	PRPE_MUSFO
23	81.5	6.4	786	1	CPOL_HPBHE
24	81	5.3	422	1	SVT2_RAT
25	81	6.3	874	1	SLAP_RAC11
26	80.5	6.3	393	1	MT04_HBFPY
27	80.5	6.3	538	1	GABE_RAT
28	80.5	6.3	732	1	NECL_RAT
29	80	6.3	354	1	R157_RUCAT
30	80	6.3	774	1	AMV2_SCHRO
31	80	6.3	953	1	YNM7_YEAST
32	79.5	6.2	312	1	TRXS_CHLTR
33	79.5	6.2	700	1	CAN2_MOUSE

31	79	6.2	504	1	YC03_KLEPN
32	78.5	6.2	860	1	SYL_ECOLI
33	78.5	6.2	887	1	THH_RAI
34	78.5	6.2	941	1	GUN_BACS6
35	78	6.1	421	1	SV11_MOUSE
36	77.5	6.1	217	1	EXPT_ERWCA
37	77.5	6.1	269	1	PENK_MOUSE
38	77.5	6.1	385	1	VGLW_YEAST
39	77.5	6.1	385	1	PSBC_CYAPA
40	77.5	6.1	461	1	ADAC_HUMAN
41	77.5	6.1	558	1	BPAC_PSEAE
42	77	6.0	373	1	ADAC_PSEAE
43	77	6.0	470	1	GLYC_CANAL
44	77	6.0	470	1	GLYC_CANAL
45	77	6.0	470	1	GLYC_CANAL

ALIGNMENTS

RESULT 1	US_GFP_ARQV2	STANDARD	ERT	238 AA
AC	24212	Q171C4		
DE	01-NOV-1995 (rel. 32, Created)			
DI	01-NOV-1995 (rel. 32, Last sequence update)			
DI	16-OCT-2001 (rel. 45, Last annotation update)			
DE	Green fluorescent protein			
GN	GFP			
OS	Aequorea victoria (Jellyfish)			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae			
OX	Aequoreidae; Aequorea			
NC	NCBI_TaxID:5100			
RN	[1]			
R2	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE			
RX	MEDLINE-92175527; PubMed-347277			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G., Cormier M.J.:			
R2	"Primary structure of the Aequorea victoria green-fluorescent protein."			
R2	Gene 111:225-233(1992).			
RN	[2]			
R2	SEQUENCE FROM N.A.			
RX	MEDLINE-94185610; PubMed-8137953			
RA	Imoye S., Isuji F.:			
R2	"Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein."			
R2	FEBS Lett. 341:277-280(1994).			
RN	[3]			
R2	CHROMOPHORE			
RX	MEDLINE-93192221; PubMed-8448132			
RA	Cody C.W., Prasher D.C., Westler W.M., Prendergast F.G., Ward W.W.:			
R2	"Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein."			
R2	Biochemistry 32:1212-1218(1993).			
RN	[4]			
R2	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS)			
RX	MEDLINE-96555565; PubMed-8703075			
RA	Ormos M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y., Remington S.J.:			
R2	"Crystal structure of the Aequorea victoria green fluorescent protein."			
R2	Science 273:1392-1395(1996).			
RN	[5]			
R2	X RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS)			
RX	MEDLINE-98254543; PubMed-9631387			
RA	Yang F., Moss L.G., Phillips G.N., Jr.:			
R2	"The molecular structure of green fluorescent protein."			
R2	Nat. Biotechnol. 14:1246-1251(1996).			
RN	[6]			
R2	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION			
RX	MEDLINE-98455509; PubMed-9782051			
RA	Wachtel R.M., Elsigier M.A., Kallio K., Hanson G.T., Remington S.J.:			
R2	"Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein."			
R2	Structure 6:1267-1277(1998).			
RN	[7]			

```

171
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS)
RP MEDLINE:99229833; PubMed:10220215;
RA Elsiger M.A., Wichter E.M., Hanson J.T., Kallio K., Nottingham S.J.
RT Structural and spectral response of green fluorescent protein
RT variants to changes in pH.
RL Biochemistry 48:5296-5301(1999).
CC -- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
CC BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN
CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
CC RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PROTEIN KINASE PKA.
CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
CC ABSORPTANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
CC PEAKS AT 503 NM WITH A SHOULDER AT 510 NM.
CC -- SUBUNIT: MONOMER.
CC -- TISSUE SPECIFICITY: PHAGOCYTES.
CC -- PIM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
CC -- BIOTECHNOLOGY: HAS become a useful and ubiquitous tool for making
CC chimeric proteins of GFP linked to other proteins where it
CC functions as a fluorescent protein tag. GFP tolerates N-and C-
CC terminal fusion to a broad variety of proteins. It has been
CC expressed in bacteria, yeast, slime mold, plants, drosophila,
CC zebrafish, and in mammalian cells. As a non-invasive fluorescent
CC marker in living cells, it allows for a wide range of applications
CC where it may function as a cell lineage tracer, reporter of gene
CC expression, or as a measure of protein-protein interactions.
CC -- DATABASE: NARS-Protein Spotlight;
CC NCPI-Issue 11 of June 2001.
CC WWW:"http://www.expasy.org/spotlight/articles/spt1011.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ac.uk/announcement/
CC or send an email to license@isb-sib.ch).
CC
EMBL MG2654; AAA27722.1;
EMBL MG2653; AAA27721.1;
EMBL L29345; AAA56246.1;
PDB 1GFL: 11-JAN-97
PDB 1EMN: 08-NOV-96
PDB 1ENB: 16-JUN-97
PDB 1EMC: 20-AUG-97
PDB 2EMC: 20-AUG-97
PDB 1EMF: 20-AUG-97
PDB 1ENF: 20-AUG-97
PDB 1EMK: 22-MAY-98
PDB 1EML: 20-AUG-97
PDB 1EMM: 20-AUG-97
PDB 2EMN: 20-AUG-97
PDB 2EMO: 20-AUG-97
PDB 1BRP: 07-JUL-97
PDB 1YFP: 26-OCT-98
PDB InterPro: IPR030786; Green_fl_protein.
PDB Pfam: PF01353; GFP.
PDB PRINTS: PRC1223; GFPLOCRESPT.
PDB ProDom: PDC13756; Green_fl_protein.
KW Luminescence; 3D-structure.
FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.
FT VARIANT 100 100 F -> Y.
FT VARIANT 108 108 T -> S.
FT VARIANT 141 141 L -> M.
FT VARIANT 219 219 V -> I.
FT CONFLICT 25 25 H -> Q (IN REF. 2).
FT CONFLICT 157 157 Q -> P (IN REF. 2).
FT CONFLICT 172 172 E -> K (IN REF. 2).
SQ SEQUENCE 238 AA; 26586 MW; EA5A6F21FBFB6E05 CRC64;

```

```

Query Match: 95.6% Score 1258; DS 1; Length 238;
Best Local Similarity 95.3%; Pred. No. 6.9e-100;
Matches 234; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

0Y : MSKSGELITGVPIIVLSDGVNCHKFSVSGRGGDATYCKZLIKFKICTGKLPVWPPTL 60
0L : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
0L 1 MSKSGELFTGVPIIVLSDGVNCHKFSVSGRGGDATYCKZLIKFKICTGKLPVWPPTL 60
0Y : VITLLTGVGVCFSRYPDHRKRDFFSKAMPDGYVOERTIFPKDWNKTKRAEYKFEETIV 120
0L : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
0L 61 VITTSYGVGVCFSRYPDHRKRDFFSKAMPDGYVOERTIFPKDGNKTKRAEYKFEETIV 120
0Y : NRIELKSGDFKEDGNILGHKLEYVNSHNHYVIMACKKNGIKVNFKIRINIEISGVQLAD 180
0L : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
0L 121 NRIELKSGDFKEDGNILGHKLEYVNSHNHYVIMACKKNGIKVNFKIRINIEISGVQLAD 180
0Y : NRIELKSGDFKEDGNILGHKLEYVNSHNHYVIMACKKNGIKVNFKIRINIEISGVQLAD 180
0L : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
0Y : HYQONTPIGCGPVLLPNNHYLSTYSALSKDPNEKRHMVLEEVTAAGTITHCMUEYK 238
0L : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
0L 161 HYQONTPIGCGPVLLPNNHYLSTYSALSKDPNEKRHMVLEEVTAAGTITHCMUEYK 238

RESULT 2
SYL_XYLFA
ID STL_XYLFA STANDARD: PRT; 879 AA.
AC 09PBG8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LEUCY-tRNA Synthetase (EC 6.1.1.4) (Leucine-tRNA ligase) (LeuRS).
DE LEUS OR XF2176.
GN Xylella fastidiosa.
OS Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella
OX NCBI_TaxID=2371;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE:20365717; PubMed=10910347;
RA SIMPSON A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.P., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.B.A., Carraro D.M., Carlier H.,
RA Colatto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C.R., Machado J.A.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.L.,
RA Fraja J.S., Franca S.C., Franco M.C., Frohme K., Furlan L.P.,
RA Gagnier M., Goldman G.H., Goldman M.B.S., Gomes S.L., Gruber A.,
RA Ho P.L., Kneisel J.D., Junqueira M.L., Kemper F.H., Kitajima J.P.,
RA Krieser J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite E.C.C.,
RA Lemos E.S.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A., Martins E.M.F., Matsukuma A.Y.,
RA Marck C.F.M., Miranda E.C., Miyaki C.Y., Monteiro-Vitorello C.E.,
RA Moon L.B., Naga M.A., Nascimento A.L.G., Netto L.F.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes G.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.S., Palmiter D.A., Paris A.,
RA Peixoto R.R., Pereira G.A.G., Pereira H.A. Jr., Pesquet C.B.,
RA Caetano R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
PA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
PA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
PA da Silveira J.F., Silvestri M.I.Z., Siqueira W.C., de Souza A.A.,
PA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
PA Vialada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
PA Zago M.A., Zatz M., Meldanis J., Setubal J.C.;
RT The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) -> AMP +
CC -!- phosphate + L-leucyl-tRNA(Leu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC

```


120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

Query Match: 7.2%; Score 92.5; DB 1; Length 533
 Best Local Similarity: 19.9%; Pred. No. 1.77
 Matches: 48; Conservative: 19; Mismatches: 97; Indels: 6; Gaps: 4

17 ELGGVNGHKFSYSGSECTATYKGL-----ELKPIK TPKKLPVPW 57
 223 EMORVLEKIGYIVNATLTGATYEVQVQAKGLNLYVCHRSTNTAGMETKSTIPWIK 282
 58 -----PTLVTLTGVCVFSYTPDMKREDFKFSAMFEGYVGEITIFPKDGNL-KTRA 110
 283 CNRFIGVDGIVETLRQAKGFEF-PETLKR-----TFEVIARELAALQDLYFNSKL 334
 111 EVK-----PEGDTLVNRETELKGLDFKEDGNILGHKLEY-----NYSKH 148
 334 QGKTACLYVGGSRSHYMNKLSKSGVDSIVAGFEFAHFKDPEVISTKIDALSKNIP 353
 149 NVYIMADKKNCKIKNFKRTHNIEDGGVGLAHYQGNTPICGHPVLLFDENHVLSTQSL 208

Query Match: 7.3%; Score 92; DB 1; Length 336
 Best Local Similarity: 24.5%; Pred. No. 3.1
 Matches: 56; Conservative: 36; Mismatches: 81; Indels: 82; Gaps: 17

19 DGVNGHKFSVSGE---GSDATYKGL-TLKFICTTKLPVPW-----PTL 60
 51 DGVDY---YNLSGELACTASIFAKIGKLDITKNFLNSGGE---WYNVVKANSEDI 142
 61 VTLLYGV-----QFSKYPDMHKR-----HDFKFSAMFEGYVGEITIFPKDGS 104
 143 KSVLTDFSYERKE-LNREP-HLRKLFEDNIYFNFSDFDENM-----MEFIGAG 192
 105 NYKTRAEVKEF-----GDTLVNRIELKGTDFKEDGNIL--GHKLEYNYSNHNVI 152
 193 NKRFELEVVEEFPKNTKSCQISNIVEINEL-IKRFD-KSMSCILIAJAHLLKENYKCY 250

394 EATVTHFQKRYVVIPEUKVESCKKAGVPLSSYGGNMKEMHDTILIDGNHHHEVWVE 453
 239 K 209
 454 K 454

RESULT 4
 YD48_METJA
 ID YD48_METJA STANDARD: PRT: 336 AA.
 AC Q58743;
 DI 15-JUN-2002 (Rel. 41, Created)
 DI 15-JUN-2002 (Rel. 41, Last sequence update)
 DI 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein KJ1348.
 GN KJ1348.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcales; Methanocaldococcus.
 CX NCBI_TaxID=2190;
 FN [1]
 PP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337599; PubMed=8686087;
 RA Bult C.J., White G., Olsen G., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Cocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Beich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gilek A.,
 RA Scott J.L., Geoghagen N.S.K., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Colton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii*;
 RL Science 273:1058-1073(1995).
 RN [2]
 PP DISCUSSION ON SEQUENCE.
 RX MEDLINE=98362732; PubMed=9697413;
 RA Kyriades N.C., Woese C.R.;
 RT *Tetrahelico-peptide-repeat proteins in the archaeon Methanococcus
 jannaschii*;
 RL Trends Biochem. Sci. 23:245-247(1998).
 CC -!- CAUTION: According to Ref.2, this sequence contains 118 repeats.
 CC These are not detected using our methodology.

Query Match: 7.1% Score 91; DB 1; Length 439;
 Post Local Similarity 20.1%; Pred No. 18; Mismatches 94; Gaps 13;
 Matches 58; Conservative 44; Mismatches 94; Gaps 13;
 QY 15 LVEDDQYNGHKFSYSSGSDATYCKLILKFICTIGKLPV-PMTFLVITITYAV----- 68
 DB 38 MNPIDTGNSTFAGVGEKND-VFKLKEKFMNLOKIPPLPPALIAIAIVSGILILIC 96
 QY 69 -----QCFSPYDPMKKHFFKSPMPSPYVOERTIFPKDSQNYKTRAEVFER----- 116
 DB 47 CIGICKKCKKKKKKKKGGK-----KNDINKK---DVKSGSSGDD 138
 QY 127 --DTLVNIELKGLDKEDGNI--LGHLEYNYNH----- 148
 DB 139 DAEIGLTGCEKEREAKKEELSKTOPSDYDFQANQITVGLIOAELPALIMGISTSPY 198
 QY 149 -NVYEMDKQKN-SIKVN-----FKIHINIEDSSVOIA-----DHVQONTIP 187
 DB 199 VKVFILPDKKKKYEKVQKKLNPFTNESFVKVPYQ-ELGKTKIMAVYFQKPSKHDC 257
 QY 186 IGUGPVLLPC-----NRYLSTQSALSKDENKKHMHVLEFFVTAAS 224
 DB 258 IGQVTVDNMTKVLGQIDFEWEDLESARKKEPKKICICTSLRYVPTAG 302
 RESULT 5
 QY62 DISOM STANDARD; PRT: 449 AA.
 AC P24506;
 DT 01-MAR-1992 (Rel. 21; Created);
 DI 01-MAR-1992 (Rel. 21; Last sequence update);
 DE 01-NOV-1995 (Rel. 32; Last annotation update);
 DE Synaptotagmin B (Synaptic vesicle protein p-65-kDa);
 GN P65-B;
 OS Discophyce omata (Electric ray);
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
 CC Elasmobranchii; Squalae; Hypnosquatae; Pristigastera; Batoida;
 CC Torpediniformes; Narcotidae; Sarcotidae; Discophyce;
 NCBI_TaxID=7785;
 [1]
 SEQUENCE FROM N.A. AND PARTIAL SEQUENCE
 MEDLINE=9127391; PubMed=2054149;
 Wordland & Miller K.G., Schilling J., Scheller R.H.;
 "Differential expression of the p65 gene family";
 Neuron 6:993-1007(1991);
 CC -1- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE MEMBRANE INTERACTIONS
 DURING TRAFFICKING OF SYNAPTIC VESICLES AT THE ACTIVE ZONE OF THE
 SYNAPSE. IT BINDS ACIDIC PHOSPHOLIPIDS WITH A SPECIFICITY THAT
 REQUIRES THE PRESENCE OF BOTH AN ACIDIC HEAD GROUP AND A DIACYL
 BACKBONE.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTRIMER (POSSIBLE).
 CC -1- SUBCELLULAR LOCATION: SYNAPTIC VESICLES IN NEURONS.
 CC -1- TISSUE SPECIFICITY: SPINAL CORD, BRAINSTEM, MIDBRAIN AND ELECTRIC
 ORGAN.
 CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed, usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcements/>
 or send an email to license@sib-sib.ch).
 CC EMBL: M64276; AAA49228.1;
 CC PIR: JH0414; JH0414;
 CC HSP: P21707; 18YN;
 CC InterPro: IPR000008; C2;
 CC InterPro: IPR002149; LRI;
 CC InterPro: IPR001565; Synaptotagmin;
 CC Pfam: PF00368; C2; 2;
 CC PRINTS: PR00360; C2DOMAIN;
 CC PRINTS: PR00390; SYNAPTOTAGM;
 CC SMART: SM00339; C2_2;
 CC PROSITE: PS00499; C2_DOMAIN_1; 2;
 CC PROSITE: PS00504; C2_DOMAIN_2; 2;
 CC Transmembrane: Repeat; Synapse; Multigene family; Glycoprotein;
 CC TRANSMEM 1 74 VESICULAR (POTENTIAL).
 CC TRANSMEM 75 101 POTENTIAL.
 CC TRANSMEM 102 439 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 153 399 PHOSPHOLIPID BINDING (PROBABLE).
 CC TRANSMEM 173 262 C2 DOMAIN 1.
 CC TRANSMEM 304 395 C2 DOMAIN 2.
 CC TRANSMEM 6 5 N-LINKED (GLYCAN).
 CC TRANSMEM 46 45 N-LINKED (GLYCAN).
 CC TRANSMEM 439 AA: 49278 MW: 203305 FMRG69F73; 49274.
 CC SEQUENCE

QY 15 LVEDDQYNGHKFSYSSGSDATYCKLILKFICTIGKLPV-PMTFLVITITYAV----- 68
 DB 38 MNPIDTGNSTFAGVGEKND-VFKLKEKFMNLOKIPPLPPALIAIAIVSGILILIC 96
 QY 69 -----QCFSPYDPMKKHFFKSPMPSPYVOERTIFPKDSQNYKTRAEVFER----- 116
 DB 47 CIGICKKCKKKKKKKKGGK-----KNDINKK---DVKSGSSGDD 138
 QY 127 --DTLVNIELKGLDKEDGNI--LGHLEYNYNH----- 148
 DB 139 DAEIGLTGCEKEREAKKEELSKTOPSDYDFQANQITVGLIOAELPALIMGISTSPY 198
 QY 149 -NVYEMDKQKN-SIKVN-----FKIHINIEDSSVOIA-----DHVQONTIP 187
 DB 199 VKVFILPDKKKKYEKVQKKLNPFTNESFVKVPYQ-ELGKTKIMAVYFQKPSKHDC 257
 QY 186 IGUGPVLLPC-----NRYLSTQSALSKDENKKHMHVLEFFVTAAS 224
 DB 258 IGQVTVDNMTKVLGQIDFEWEDLESARKKEPKKICICTSLRYVPTAG 302
 RESULT 6
 QY62 HUMAN STANDARD; PRT: 1224 AA.
 ID COPALHUMAN
 AC P58521;
 DT 01-OCT-1996 (Rel. 34; Created);
 DI 01-OCT-1996 (Rel. 34; Last sequence update);
 DT 15-JUN-2002 (Rel. 41; Last annotation update);
 DE Coatomer alpha subunit (Alpha-coat protein) (Alpha-COP) (HEPCOP)
 DE (HEPCOP) [contains: Xenin (Xenopsin-related peptide); Proxenin].
 GN COPA.
 OS Homo sapiens (Human);
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95194806; PubMed=8647451;
 RA Chow V.T.K., Ques H.H.;
 RT "HEP-COP, a novel human gene whose product is highly homologous to
 the alpha-subunit of the yeast coatomer protein complex";
 KL Gene 165:223-227(1996);
 RN [2]
 RP SEQUENCE OF 1-25 (XENIN).
 RC TISSUE=gastric mucosa;
 RX MEDLINE=93054515; PubMed=1429581;
 RA Feurle G.E., Hamscher G., Kusiek R., Meyer H.E., Metzger J.W.;
 RT "Identification of xenin, a xenopsin-related peptide, in the human
 gastric mucosa and its effect on exocrine pancreatic secretion";
 KL J. Biol. Chem. 267:22305-22309(1992);
 RN [3]
 RP PROCESSING OF COPA TO PRODUCE XENIN.
 RX MEDLINE=98032493; PubMed=9365789;
 RA Chow V.T., Ques H.H.;
 RT "Alpha coat protein COPA (HEP-COP): presence of an Alu repeat in cDNA
 and identity of the amino terminus to xenin";
 KL Ann. Hum. Genet. 51:369-373(1997);
 RN [4]
 RP REVIEW ON XENIN.
 RX MEDLINE=98192336; PubMed=9533652;
 RA Feurle G.E.;
 RT "Xenin: a review";
 KL Peptides 19:609-615(1998).
 CC -1- FUNCTION: THE COATOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
 TO D-LYSINE MOTIFS AND REVERSIBLY ASSOCIATES WITH GOLGI NON-
 CLATHRIN-COATED VESICLES, WHICH FURTHER MEDiate BIOSYNTHETIC
 PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI UP TO THE TRANS GOLGI
 NETWORK. COATOMER COMPLEX IS REQUIRED FOR BUDDING FROM GOLGI
 MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GOLGI-TO-ER

DR SMART: SMO0486; POLBO: 1.
 DR PROSITE: PS00116; DNA-POLYMERASE; H: FA-SE-NES;
 KW Transerase: DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding Zinc-finger; Nuclear protein;
 FT ZNFING 2108 2191 POTENTIAL;
 FT VARIANT 544 544 M + S 1 (IN POL2-16 TS MUTANT);
 FT VARIANT 710 710 P + S 8 (IN POL2-16 TS MUTANT);
 SQ SEQUENCE 2222 AA: 255665 MW: 6870222341471658 kDa

Query Match 6.48; Score 86.5; DB 1; Length 2222;
 Best Local Similarity 28.23; Pred. No. 11;
 Matches 37; Conservative 14; Mismatches 40; Indels 30; Gaps 71

QY 53 LPVWPV-TLVITTYGVQVTSRYTHUM -----KRDPSKSAKEGVV-- QERTY 98
 Db 983 LPSFPEPTYPELENGAKKLYSLVPCSLMYRVKPTNRWDELKIDLVNLYVETHEHTI 942
 QY 99 PFKDDGNYKTR-AEVFEEDLVNLR-----IELKIDFKEDGNILGHKLYNTN 146
 Db 943 FFEVDGYKAMILPSSKEEGKSKRYAVENEDSSLAELGPELKERGEL--QLKNEQ 989
 QY 147 S-HNVVIMAD 255
 Db 100C SDPKVFEEDG 1010

RESULT 10
 ID COPA_BOVIN STANDARD: PRI: 1224 AA.
 AC Q27954;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 43, Last annotation update)
 DE Coatmer alpha subunit (Alpha-coat protein) (Alpha-COP) (HEP-COP)
 DE (HEP-COP) [Contains: Xenin (Xenopus)-related peptide]; Proxenin;
 GN COPA.
 CS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 PN 11;
 RP SEQUENCE FROM N.A.
 RA Weland F.11;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: (MAR-1996) THE COATMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
 CC TO GLYCINE MOTIFS AND REVERSIBLY ASSOCIATES WITH GOLGI MEM-
 CC BRANE-RELATED VESICLES, WHICH FURTHER MEDIATES BIOSYNTHETIC
 CC PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI, TO THE TRANS GOLGI;
 CC NETWORK. COATMER COMPLEX IS REQUIRED FOR BUILDING FROM GOLGI
 CC MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GLYT-TO-ER
 CC TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COATMER
 CC CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ANP-R-PROSYLATION
 CC FACTORS (AREFS), WHICH ARE SMALL GTP-BINDING PROTEINS; THE COMPLEX
 CC ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
 CC PROCESSING, ACTIVITY, AND ENDOCYTIC RECYCLING OF LIL RECEPTORS (BY
 CC SIMILARITY).
 CC -!- FUNCTION: XENIN STIMULATES EXOCRINE PANCREATIC SECRETION. IT
 CC INHIBITS PENTAGASTRIN-STIMULATED SECRETION OF ACID, TO INHIBIT
 CC EXOCRINE PANCREATIC SECRETION AND TO AFFECT SMALL AND LARGE
 CC INTESTINAL MOTILITY. IN THE GUT, XENIN INTERACTS WITH THE
 CC NEUTENSIN RECEPTOR (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
 CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: THE COATMER IS CYTOSOLIC OR POLYMERIZED
 CC ON THE CYTOSOLIC SIDE OF THE GOLGI, AS WELL AS IN THE
 CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY). XENIN IS
 CC SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP (=MANS)).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, EMBL-EBI
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi.ac.uk/announce/>
 CC or send an email to license@ebi.ac.uk.)

DR EMBL: X26768; CA65543.1;
 DR InterPro: IPR001060; WD46;
 DR Pfam: PF00400; WD46.7;
 DR PRINTS: PR00320; GPR07F1NSPT.
 DR PRODOM: PR000028; WD46.7;
 DR SMART: SMO0320; WD46.7;
 DR PROSITE: PS00678; WD REPEATS_1; 2;
 DR PROSITE: PS00882; WD REPEATS_2; 6;
 DR PROSITE: PS00294; WD REPEATS_REGION; 1;
 KW Transport: Protein transport; Golgi stack; Endoplasmic reticulum;
 KW Membrane: Repeat; WD repeat; Hormone.
 FT PEPTIDE 1 25 XENIN.
 FT PEPTIDE 1 35 PROXENIN.
 FT REPEAT 7 37 WD 1.
 FT REPEAT 49 79 WD 2.
 FT REPEAT 91 121 WD 3.
 FT REPEAT 133 163 WD 4.
 FT REPEAT 203 233 WD 5.
 FT REPEAT 247 277 WD 6.
 SQ SEQUENCE 1224 AA: 138358 MW: 693508292006830 kDa

Query Match 6.78; Score 86; DB 1; Length 1224;
 Best Local Similarity 25.38; Pred. No. 16;
 Matches 40; Conservative 27; Mismatches 59; Indels 32; Gaps 71

QY 90 EGYVOERTIFKDDGNKYTRAEVFKPGDTVNRIR-LKGDIFKEDGNILGHKLYNTN 148
 Db 632 KGYPEVALHFVKDE---KTRPSLAECGNCIEALEAAKALDDKNCKEKLGFVALLQSNHQ 688
 QY 149 NVYIMADKQNGIKVNF-----KIRHNIEDGGVQ--LADHYQONTPTGDPVLLPO 197
 Db 689 -VEMYQYQTKNFDKLSFLYLITGNLEKJRKMKKIAE-RKMSGHYONALYGD----- 741
 QY 198 NHYLSTQSALSKDPNKKRDMVLEFVTAAGITHGMD 235
 Db 742 ---VSEVRVRLKNCGCKS-----LAYLFAA--THGLPE 769

RESULT 11
 ID CAN2_HUMAN STANDARD: PRI: 700 AA.
 AC P17555; O14736;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calpain 2, large (catalytic) subunit precursor (EC 3.4.22.17)
 DE (Calcium-activated neutral proteinase) (CANP) (M-type) (N-calpain)
 DE (Milimolar-calpain).
 GN CAPN2 OF CAPN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 PN 11;
 RP SEQUENCE FROM N.A.
 RA MEDLINE:89166474; PubMed:2852952;
 RA Itajoh S., Aoki K., Ohno S., Emori Y., Kawasaki H., Suqihara H.,
 RA Suzuki K.;
 RA "Molecular cloning of the cDNA for the large subunit of the
 RA high-Ca2+-requiring form of human Ca2+-activated neutral proteinase";
 RA Biochemistry 27:8122-8128(1988).
 RN [2]
 RP SEQUENCE OF 1-79 FROM N.A.
 RC TISSUE:lymph node;
 RX MEDLINE:89197947; PubMed:2519381;
 RX Hata A., Ohno S., Akita Y., Suzuki K.;
 RX "Tandemly reiterated negative enhancer-like elements regulate
 RX transcription of a human gene for the large subunit of calcium-

dependent protease.";
 C. Biol. Chem. 264:6404-6411(1989).
 [3]
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RP MEDLINE=22105516; PubMed=10639123;
 RA Stroth, S., Fernandez-Catalan C., Braun M., Haber P., Masumoto H.,
 RA Nakagawa K., Irie A., Sorimachi H., Bourenkov G., Bartunik H.,
 RA Suzuki K., Hode W.;
 RT "The crystal structure of calcium-free human metalloprotease suggests an
 RT electrostatic switch mechanism for activation by calcium."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr- Xaa, Met- Xaa or
 CC Arg-1-Xaa with Leu or Val as the P2 residue.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding.
 CC -!- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM BINDING DOMAINS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 02.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M23254; AAA35645.1;
 DR EMBL: J04706; AAA52760.1;
 DR PIR: A31218; A31218;
 DR PDR: 1KFU; C7-DEP-G;
 DR MEROPS: C02.002;
 DR Genbank: HGNC:1475; CAI2;
 DR MIM: 114230;
 DR InterPro: IPR002548; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR003169; SHP1-related.
 DR Pfam: PF00036; ef-hand_3.
 DR Pfam: PF00648; Peptidase_C2_1.
 DR Pfam: PF01067; Calpain_1; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00230; Cyspe; 1.
 DR SMART: SM00054; EFh; 2.
 DR PROSITE: PS00018; EF-HAND; 2.
 DR PROSITE: PS00139; THIOL-PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL-PROTEASE_HIS; FALSE-NEG.
 DR PROSITE: PS00640; THIOL-PROTEASE_ASN; FALSE-NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Repeat: Multiple family;
 3D-structure.
 FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
 FT CHAIN 20 790 CALPAIN 2, LARGE (CATALYTIC) SUBUNIT.
 FT DOMAIN 20 210 THIOL-PROTEASE DOMAIN 1.
 FT DOMAIN 211 395 THIOL-PROTEASE DOMAIN 1.
 FT DOMAIN 355 514 THIOL-PROTEASE DOMAIN 1.
 FT DOMAIN 515 523 ZINKER.
 FT DOMAIN 530 700 DOMAIN IV.
 FT DOMAIN 541 552 EF-HAND 1.
 FT CALBIND 585 596 EF-HAND 2.
 FT CALBIND 615 626 EF-HAND 3.
 FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 175 165 BY SIMILARITY.
 FT ACT_SITE 262 262 BY SIMILARITY.
 FT ACT_SITE 286 286 BY SIMILARITY.
 FT CONFLICT 68 S 55 G (IN REF. 2).

EC CONELICI 73 74 MR -> IE (IN REF. 2)
 SO SEQUENCE 700 AA: 80006 MW: A944D13BC8465531 CRC64;
 Query Match 6.7% Score 85.5; DB 1: Length 700;
 Best Local Similarity 20.9%; Pred. No. 9.2;
 Matches 47: Conservative 39; Mismatches 80; Indels 59; Gaps 10;
 QY 4 GBEIET---GVYPIVLELDGVNHHK---FSVSGEGGDATYKLTILKFKICT--GKLPVP 55
 DB 423 GEDMHTIGFIVPEELSGQTNHLSKNFELNKHAKERSDTF--INLREVNREKLP-P 479
 QY 57 WPIYLTITLYG-----VQCFS-RYEDHMKRHCFFKSAPEGVVQKGTFFEDGQNYK 107
 DB 480 GSYIVPSTFFPKKGDGFCIRVFSEKKADYQAVDDIEANLEEDFISQDDI---QDVVR 536
 QY 106 TRAFV-----KFEQJLVNRIELKQIDKNDG----- 134
 DB 517 LPAQLAGSAEIASAPELOILRRVLAKRQDIKDSFETICKIMVDM:JNSGSKLGLKE 596
 QY 135 -NLQHKLFYNNSHNYIMADKQKNG:KVNFKIHUNIED:GVOL 178
 DB 507 FVILWKIG---KYQKIVAEIDVDRSGTNSYEMKKALEHAGPKM 639
 RESULT 12
 Q152_HABIN STANDARD PRI: 795 AA.
 AC P44955;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 ID D15-Ag (Outer membrane protein D15).
 GN H109.7
 OS Haemophilus influenzae
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales
 OC Haemophilus
 OX NCBI_TaxID:727;
 RN 1;
 RP SEQUENCE FROM N.A.
 RC STRAIN=BJ / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed:7542866;
 RA Frieschmann R.D., Adams M.B., White G., Clayton P.A., Kirkness E.F.,
 RA Korlavage A.R., Buit C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Bink L.L., Gladek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.B.,
 RA Jeterlack J.R., Hanna M.C., Nguyen D.T., Saudek D.W., Brandes R.C.,
 RA Fine L.S., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
 RA Greth C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.D.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U32773; AAC22575.1;
 DR FICR: H10517;
 DR InterPro: IPR000184; Bac_surfac_D15.
 DR Pfam: PF01103; Bac_surfac_Ag; 1.
 KW Antigen; Outer membrane; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
 FT SEQUENCE 795 AA: 87478 MW: B85691FC22E6ED44 CRC64;

```

Query Match      6.78; Score 45.5; DB 1; Length 797;
Post-Score Similarity 42.4%; Pred. No. 11;
Matches 49; Conservative 39; Mismatches 72; Indels 43; Gaps 12;

66 YGVGCFKSYDHHKRMLEF-----FKSAMPEFYVDE-----RTIFF 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 YGIFSGISYCAWVKQNFLEIGAAVSINIKNIYUTSVNCAJTEYFTKGVSLGQVFF 488
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 KDCQNYIRAEVFEHETLVNRIELKGIQFKEGGNI-----LGH-----KLENYNSHN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 ENYNSKSDTSYNYKPTTKGSNVLTGFQVNNNSYVGLGHYNYKISFEALFN---RN 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 VYEMADQKNGKVN-----EKIRN-----LEDGGVQLADHYQQNTVCGGVVLP- 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
545 LYIOSMKFENGKTKNDPDEFTGWNINSNLRGYFTPKGVKAS-----LSGVETIG 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

197 -DNVYSTQSAISKQNEKROHMVLEFPVTAQIG 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
596 SNNKYKLSAEVOGFYPLDRDHLWVYSAKASAGYANG 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULTS 14
ID CAN2_PIG STANDARD: PRT: 324 AA.
AC P43367:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-JUN-2002 (Rel. 41, Last annotation update)
DE Calpain-2, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
DE neutral proteinase) (CAMP) (M-type) (M-calpain) (Millimolar-calpain)
DE (Fragment).
DE CAPN2.
OS Sus. scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9423:
XX [1]
RN SEQUENCE OF 3-204 FROM N.A.
RP TISSUE=Skeletal muscle.
RC MEDLINE 94146155; PubMed=212396:
RX Sun W., Ji S.-Q., Ebert P.J., Bidwell C.A., Hancock J.B.:
RA "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine
RT skeletal muscle.".
RL Biochimie 75:931-935(1993).
RN [2]
RP SEQUENCE OF 122-224 FROM N.A.
RC TISSUE=Pulmonary artery.
RX MEDLINE=98405921; PubMed=9728240:
RA Zhang J.L., Patel J.M., Block F.R.:
RT "Hypoxia-specific upregulation of calpain activity and gene
RE expression in pulmonary artery endothelial cells.".
Am. J. Physiol. 275:H461-H468(1998).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction (By similarity).
CC -2- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
CC Arg-I-Xaa with Leu or Val as the P2 residue.
CC -3- COFACTOR: Binds 3 calcium ions.
CC -4- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC of calcium and inhibited by calpastatin.
CC -5- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -6- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding.
CC -7- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
CC -8- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities without the written permission of the Swiss Institute of
CC Bioinformatics is prohibited.

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)
 CC or send an email to license@isb-sib.ch.

CC EMBL: U01182; AAC48401.1; ..
 CC EMBL: U01320; AAB17391.1; ..
 CC DR HSP: Q07009; IDP0.
 CC DR MEROPS: C02.902; ..
 CC DR InterPro: IP0002048; EF-hand.
 CC DR InterPro: IP0001300; Protease_C2.
 CC DR InterPro: IP000169; Spro_CasSite.
 CC DR Pfam: PF00036; ehand; 2.
 CC DR Pfam: PF01567; Calpain_1; 1.
 CC DR ProDom: PD00012; EF-hand; 1.
 CC DR ProSITE: PS00018; EF-HAND; 1.
 CC DR ProSITE: PS00139; THIOL_PROTEASE_CYS; PARTIAL.
 CC DR ProSITE: PS00639; THIOL_PROTEASE_HIS; PARTIAL.
 CC DR ProSITE: PS00640; THIOL_PROTEASE_ASN; PARTIAL.
 CC KW HydroLase; Thiol protease; Calcium-binding; Repeat; Multigene family.
 CC FT NON-TER 1 157 ..
 CC FT DOMAIN <1 157 ..
 CC FT DOMAIN 158 324 ..
 CC FT DOMAIN <1 138 ..
 CC FT DOMAIN 139 153 ..
 CC FT DOMAIN 154 324 ..
 CC FT CALBIND 165 176 ..
 CC FT CALBIND 209 220 ..
 CC FT CALBIND 239 250 ..
 CC FT CALBIND 274 285 ..
 CC FT DOMAIN 304 315 ..
 CC FT DOMAIN 304 315 ..
 CC FT CONFLICT 202 202 ..
 CC SQ SEQUENCE 324 AA; 37808 MW; 3925555245E1230F CDS64;
 CC Query Match 6.6%; Score 84.5; DB 1; Length 324;
 CC Best Local Similarity 19.6%; Pred. No. 4.4;
 CC Matches 55; Conservative 49; Mismatches 139; Indels 67; Gaps 120
 CC
 CC QY 4 QUESLEP---GVVPLVLEGGVNGHK-----PSVDSGSHATYSKIILKKECTTKLPWP 56
 CC DB GEDMTIGGGIYEVEPELEGGVNGHK-----PSVDSGSHATYSKIILKKECTTKLPWP 56
 CC QY 47 GEDMTIGGGIYEVEPELEGGVNGHK-----PSVDSGSHATYSKIILKKECTTKLPWP 109
 CC DB GEDMTIGGGIYEVEPELEGGVNGHK-----PSVDSGSHATYSKIILKKECTTKLPWP 109
 CC QY 57 KPTLVITLYG-----VQCFE-KYDPMKRRHFFKSAHPSSVQVEELFFKDCGNK 107
 CC DB KPTLVITLYG-----VQCFE-KYDPMKRRHFFKSAHPSSVQVEELFFKDCGNK 107
 CC QY 104 GYVLPVPTFPPNKGDDGCIIVFKEKKKQYGVVDELPADDEENTASEDEI---LDGPR 160
 CC DB GYVLPVPTFPPNKGDDGCIIVFKEKKKQYGVVDELPADDEENTASEDEI---LDGPR 160
 CC QY 138 TRAEV-----KPEGTPLVNLRIELKQIDFEKAS----- 134
 CC DB TRAEV-----KPEGTPLVNLRIELKQIDFEKAS----- 134
 CC QY 141 LFAQLAGDAELISAPELDTILSLVLAKQDICKSDTFSEIEELKLVVWVDSGSAKLHKE 220
 CC DB LFAQLAGDAELISAPELDTILSLVLAKQDICKSDTFSEIEELKLVVWVDSGSAKLHKE 220
 CC QY 135 NLLQHKLEYNYSNIVYIMAKCKNGIKVFKLHNTHDQVVALHYQG--NTPIDGG 191
 CC DB NLLQHKLEYNYSNIVYIMAKCKNGIKVFKLHNTHDQVVALHYQG--NTPIDGG 191
 CC TB 221 FYILWKTKIO---KYQKIYRELDVDSGTPNSYKPKALEEMFKELQTHQVAVAFADG 277
 CC QY 192 PVLLPENNIV-----LSTQSALSKDPNEKROHMVILDEVT 426
 CC DB PVLLPENNIV-----LSTQSALSKDPNEKROHMVILDEVT 426
 CC QY 278 QLIIDFQNFVGLVLELEFRISKOLDSENTCTFLDLAIS 317
 CC DB QLIIDFQNFVGLVLELEFRISKOLDSENTCTFLDLAIS 317
 CC
 CC RESULT 15
 CC D153.HAEIN
 CC ID D153.HAEIN STANDARD: PRI: 793 AA.
 CC AC Q32629.
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Protective surface antigen D15 precursor (85 kDa D15 antigen)
 CC DE (p-15-Ag) (Outer membrane protein I15).
 CC OS Haemophilus influenzae
 CC CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales
 CC CC Haemophilus
 CC CC NCBI_TaxID:727
 CC RN
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-PAK 12085;

RX MEDLINE:37427952; PubMed:9294140;
 RA Loomis S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
 KA Klein M.H.;
 RT Outer membrane protein D15 is conserved among Haemophilus influenzae
 KI species and may represent a universal protective antigen against
 RI invasive disease.
 RL Infect. Immun. 65:3701-3707(1997).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U06934; AAB01977.1; ..
 CC InterPro: IP000184; Bac_surfAg_D15.
 CC Pfam: PF01103; Bac_surfAg; 1.
 CC KW Antigen; Outer membrane; Signal.
 CC FT SIGNAL 1 19 POTENTIAL.
 CC FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
 CC SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2035801A14 CDS64;
 CC
 CC Query Match 6.6%; Score 84.5; DB 1; Length 793;
 CC Best Local Similarity 23.0%; Pred. No. 13;
 CC Matches 50; Conservative 31; Mismatches 73; Indels 6; Gaps 12;
 CC
 CC QY 56 YGVQCFEYRYPDHMKRDF-----FKSAMPEGVQE-----RIFF 100
 CC DB YGVQCFEYRYPDHMKRDF-----FKSAMPEGVQE-----RIFF 100
 CC QY 429 YGVQCFEYRYPDHMKRDF-----FKSAMPEGVQE-----RIFF 488
 CC DB YGVQCFEYRYPDHMKRDF-----FKSAMPEGVQE-----RIFF 488
 CC QY 101 KQSGNYKTRAEYKPEPDLVNLRIELKQIDFEKAS-----LGR-----KLFYNNSHN 149
 CC DB KQSGNYKTRAEYKPEPDLVNLRIELKQIDFEKAS-----LGR-----KLFYNNSHN 149
 CC QY 489 EVNDSKSDTSSSKYKATYGSNVTG-CFVNNNSYVVGIGHYNNKISNPALEYN---RN 544
 CC DB EVNDSKSDTSSSKYKATYGSNVTG-CFVNNNSYVVGIGHYNNKISNPALEYN---RN 544
 CC QY 150 VYIMAEKQK-NGIKVN-----FKIRHN-----IDGGVQIADHYQQNTPIDGGVLLP- 156
 CC DB VYIMAEKQK-NGIKVN-----FKIRHN-----IDGGVQIADHYQQNTPIDGGVLLP- 156
 CC QY 545 LYIQSMKFKGNGIKTNEFDESGNNYNSLNRGYFFTKGVKAS-----LS-GRVTIPS 595
 CC DB LYIQSMKFKGNGIKTNEFDESGNNYNSLNRGYFFTKGVKAS-----LS-GRVTIPS 595
 CC QY 197 -DNHLYLSTQSALSKDPNEKROHMVILDEVTAAAGIHG 242
 CC DB -DNHLYLSTQSALSKDPNEKROHMVILDEVTAAAGIHG 242
 CC QY 596 SDNKYKRLSACVQGYPLDRDHRHWVVSASAKASYANG 632
 CC DB SDNKYKRLSACVQGYPLDRDHRHWVVSASAKASYANG 632
 CC
 CC Search completed: July 16, 2003, 16:11:19
 CC Job time : 11.5667 secs

GeneCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.33s Seconds
(without alignments)
1166.433 Million cell updates/sec

Title: us-09-967-301-4

Perfect score: 1276
Sequence: 1 MSKGESEFGVPIIVELDG.....VLEFVTAAGTPEGKFLYR 248

Scoring table: RUOSUM62
Gapop 10.0 ; Gapext 6.5

Searched: 671586 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671586

Minimum DB seq length: 6
Maximum DB seq length: 200000000

Post-processing: Minimum Match: 04
Maximum Match: 2034
Listing: 11rst: 45 summaries

Database:

SPTREML_21:

1: sp-archaeal*
2: sp-bacteria*
3: sp-fungi*
4: sp-human*
5: sp-invertebrates*
6: sp-mammal*
7: sp-misc*
8: sp-ornamental*
9: sp-plaqua*
10: sp-plant*
11: sp-rodent*
12: sp-virus*
13: sp-vertebrate*
14: sp-unclassified*
15: sp-virus*
16: sp-bacterioph*
17: sp-archaeal*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result NO.	Score	Query Match	Length	DB ID	Description
1	1258	98.6	238	5 Q27953	Q27953 unidentified
2	1252	98.1	238	5 Q93125	Q93125 aequorea vi
3	1216	95.3	238	5 Q17105	Q17105 aequorea vi
4	1215	95.2	758	4 Q48001	Q48001 humo sapie
5	1201	94.1	238	5 Q17106	Q17106 aequorea vi
6	1109	86.9	238	5 Q8WTC6	Q8WTC6 aequorea ma
7	1105	86.6	238	5 Q8WPC5	Q8WPC5 aequorea ma
8	1102	86.4	238	5 Q8WTC4	Q8WTC4 aequorea ma
9	1100	86.2	238	5 Q8WTD0	Q8WTD0 aequorea ma
10	1099	86.1	238	5 Q8WTC9	Q8WTC9 aequorea ma
11	1099	86.1	238	5 Q8WTC8	Q8WTC8 aequorea ma
12	1095	85.9	238	5 Q8WTC7	Q8WTC7 aequorea ma
13	1095	85.8	238	5 Q8WTC5	Q8WTC5 aequorea ma
14	251.5	19.7	225	5 Q950A7	Q950A7 montastraea
15	243.5	19.6	236	5 Q8T6D0	Q8T6D0 dendronephr
16	249	19.5	225	5 Q933F5	Q933F5 montastraea

17	243.5	19.1	266	5 Q9U6R3	Q9U6R3 clavularia
18	236.5	16.5	225	5 Q8T5F1	Q8T5F1 montastraea
19	210.5	16.5	234	5 Q8U5F2	Q8U5F2 montastraea
20	210	16.5	229	5 Q9U6V6	Q9U6V6 anemonia ma
21	209.5	16.4	225	5 Q9U6V9	Q9U6V9 discosoria s
22	209	16.4	221	5 Q95PC4	Q95PC4 conioptora t
23	208	16.3	227	5 Q962P9	Q962P9 montastraea
24	205.5	16.1	232	5 Q9GPT5	Q9GPT5 anemonia su
25	204.5	16.0	235	5 Q8T5F0	Q8T5F0 scolymia cu
26	204	16.0	227	5 Q95VT0	Q95VT0 montastraea
27	204	16.0	238	5 Q9BLV9	Q9BLV9 renilla mue
28	203	15.9	232	5 Q9U6V7	Q9U6V7 discosoria s
29	201.5	15.8	232	5 Q932Z8	Q932Z8 anemonia su
30	197.5	15.5	225	5 Q8T6T9	Q8T6T9 heteractis
31	195	15.3	230	5 Q9GJ37	Q9GJ37 discosoma s
32	187.5	14.7	227	5 Q95W85	Q95W85 heteractis
33	187	14.7	228	5 Q9GB16	Q9GB16 anemonia su
34	182.5	14.3	231	5 Q9U6V5	Q9U6V5 zoanthus sp
35	181.5	14.2	238	5 Q9HJ20	Q9HJ20 plilosarcus
36	181	14.2	233	5 Q96319	Q96319 renilla ren
37	179.5	14.1	227	5 Q95W95	Q95W95 condylactis
38	175	13.7	229	5 Q8T5B7	Q8T5B7 condylactis
39	173.5	13.6	227	5 Q95W11	Q95W11 condylactis
40	173.5	13.6	231	5 Q8T5B8	Q8T5B8 riordea fl
41	171	13.4	231	5 Q9UGV4	Q9UGV4 zoanthus sp
42	166.5	13.0	231	5 Q8T4M4	Q8T4M4 zoanthus sp
43	165	12.9	234	5 Q8T5F3	Q8T5F3 scolymia cu
44	152.5	12.0	231	5 Q8T6T8	Q8T6T8 discosoma s
45	142	11.1	231	5 Q8T5E9	Q8T5E9 riordea fl

ALIGNMENTS

RESULT:

Q27953 PRELIMINARY: PR: 238 AA.
AC Q27953
CT 01-NOV-1996 (TRENBLER, C.: Created)
LI 01-NOV-1996 (TRENBLER, C.: Last sequence update)
PT 01-JUN-2002 (TRENBLER, C.: Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI:TaxID=12644:
RN 111
R1 SEQUENCE FROM N.A.
R2 MEDLINE-97299832; PubMed-9154981;
R3 Kowandall G.J.A., Mendes G., Wolbert E.J.H., de Boer A.B.:
RT "Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage";
RL Plant Mol Biol 41:889-903(1997).
DR EMBL: X96418; CAA65278.1;
DR RSP: P42212; IGFL.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF03533; GFP.
DR PRINTS: PR01229; GFP-UNDESCENT.
DR ProDom: PD018756; Green_fl_protein; 1.
SQ SEQUENCE: 238 AA: 25884 MW: CA932D47262AF2D3 CRC64;

Query Match 98.6% Score 1258; DB 5; Length 238;
Best Local Similarity 98.3% Pred. No. 8.2e-98;
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVPIIVELDGNNHKSFGSGEGSDATYKGLTKLFCTTGKLPVWPIL 60
DE 1 MKRGEELFGVPIIVELDGNNHKSFGSGEGSDATYKGLTKLFCTTGKLPVWPIL 60
QY 61 VTITLVGQCFSRYPDMHMRHDFKFSAMPGYVCGERTIFPKDGNKTRAEVKEFGTLY 120
DE 61 VTITLVGQCFSRYPDMHMRHDFKFSAMPGYVCGERTIFPKDGNKTRAEVKEFGTLY 120


```

SQ SEQUENCE 758 AA: 85015 MW: 8612408660709D49 CRC64:
Query Match 95.2%; Score 12.5; DB 5; Length 238;
Best Local Similarity 95.6%; Pred. No. 1.5e-93;
Matches 222; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 60
DB 499 VSKGESLFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 558
QY 61 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 120
DB 499 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 612
QY 121 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
DB 499 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238
DB 499 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238

RESULT 5
Q17106 PRELIMINARY: PRT: 238 AA.
AC Q17106
QY 01-NOV-1996 (TrEMBLrel. 01, Created)
DB 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN 1
RP SEQUENCE FROM N.A.
RA Watkins J.K., Campbell A.K.,
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X83960; CAA58790.1;
DK HSSP: P42222; IRFP.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFPLORESCENT.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TPS 238
SQ SEQUENCE 238 AA: 26867 MW: 80464626208A8U4 CRC64:

Query Match 94.1%; Score 12.0; DB 5; Length 238;
Best Local Similarity 93.3%; Pred. No. 5e-93;
Matches 222; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 60
DB 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 60
QY 61 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 120
DB 61 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 120
QY 121 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
DB 121 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238

RESULT 6
Q8WTC6 PRELIMINARY: PRT: 238 AA.
AC Q8WTC6
QY 01-MAR-2002 (TrEMBLrel. 20, Created)
DB 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactylia.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qiu Y.X., Pang S.Q.,
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013821; AAK02059.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TPS 238
SQ SEQUENCE 238 AA: 27049 MW: 8185005529012B CRC64:

Query Match 86.6%; Score 11.05; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 5.6e-95;

```

```

AC Q8WTC6
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactylia.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qiu Y.X., Pang S.Q.,
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435431; AA53316.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TPS 238
SQ SEQUENCE 238 AA: 27015 MW: 658ED75E88926993 CRC64:

Query Match 86.3%; Score 11.59; DB 5; Length 238;
Best Local Similarity 83.8%; Pred. No. 2.6e-95;
Matches 199; Conservative 20; Mismatches 19; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 60
DB 1 MSKGEELFTGVVPLVLDGVDVNGHKFVSQSGEGDATTGKLTIKFICTGKLPVWPTL 60
QY 61 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 120
DB 61 VTITSYGVQCFSPRYPDHMKHDFKSAPEGYVOERTIFFKDGNGKIRAEVKKFECTIV 120
QY 121 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
DB 121 NRLEKLGIDFKEDGNILGHKLEYNYSNHNYYIMADKQKNGIKVNEKIRHNIEGSGVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKPNEKRHMVLEFVTAAGTTHGMDELYK 238

RESULT 7
Q8RP95 PRELIMINARY: PRT: 238 AA.
AC Q8RP95
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPXM.
OS Aequorea macrodactylia.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN 1
RP SEQUENCE FROM N.A.
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qiu Y.X., Pang S.Q.,
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013824; AAK02052.1;
DK InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
FT NON_TPS 238
SQ SEQUENCE 238 AA: 27049 MW: 8185005529012B CRC64:

Query Match 86.6%; Score 11.05; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 5.6e-95;

```

```

Matches 198: Conservative 20; Mismatches 20; Indels 0; Gaps 0
QY 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
QY 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
DB 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238

RESULT 8
Q8WTC4 PRELIMINARY: PRT: 238 AA.
ID Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9 PRELIMINARY: PRT: 238 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DI 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHG24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL33913;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF70E5A73A CRC64;

Query Match 86.4%; Score 1102; DB 5; Length 238;
Best Local Similarity 84.5%; Pred. No. 1.8e-84;
Matches 201; Conservative 16; Mismatches 20; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
QY 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
DB 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238

RESULT 9
Q8WTC0 PRELIMINARY: PRT: 238 AA.
ID Q8WTC0 PRELIMINARY: PRT: 238 AA.
AC Q8WTC0 PRELIMINARY: PRT: 238 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DI 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHG24;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL33913;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26956 MW; 75521EAF70E5A73A CRC64;

Query Match 86.4%; Score 1102; DB 5; Length 238;
Best Local Similarity 84.5%; Pred. No. 1.8e-84;
Matches 201; Conservative 16; Mismatches 20; Indels 0; Gaps 0;

```

```

LT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GFPXMI161;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435427; AAL33913;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26997 MW; 5F80A192174CP64D CRC64;

Query Match 86.2%; Score 1100; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 1.5e-84;
Matches 198; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYKGLKFTCTTCKLPVWPPTL 60
QY 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
DB 61 VTTLYGVQGSFSPYDMHKKHDFPKSAMPEGYVOERTIFFKDDGNYKTRAEVKESETLV 120
QY 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
DB 121 NRIELKGIDFKEGNILGHKLEYNNSHNYIMADKNGIKYKFKIRHNIEDGGVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTALSADKPNKRDHMLLEFVTAAGTTHSKMDELYK 238

RESULT 10
Q8WTC5 PRELIMINARY: PRT: 238 AA.
ID Q8WTC9 PRELIMINARY: PRT: 238 AA.
AC Q8WTC9 PRELIMINARY: PRT: 238 AA.
DT 01-MAR-2002 (Tremblrel. 20, Created)
DI 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GFPXMI162;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435428; AAL33913;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;

Query Match 86.1%; Score 1099; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 1.8e-84;
Matches 198; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

```



```

DB 1 MSKSELEFGVVPVILVELDGVGHGKFSVSGEGEDADYVGRLEIRFICITLKLIVLWPTL 60
QY 61 VTLLYGVQCSRYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
DB 62 VTLLYGVQCSRYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
QY 63 VTLLYGVQCSRYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
DB 64 VTLLYGVQCSRYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
QY 121 NRIELKGLDQFKEGNTAGKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
DB 122 NRIELKGLDQFKEGNTAGKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
QY 123 NRIELKGLDQFKEGNTAGKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
DB 124 NRIELKGLDQFKEGNTAGKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
QY 161 HYQQTNPICGLGVLPENNYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
DB 162 HYQQTNPICGLGVLPENNYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
QY 181 HYQQTNPICGLGVLPENNYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
DB 182 HYQQTNPICGLGVLPENNYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240

RESULT 14
Q950A7 PRELIMINARY: PRT: 225 AA.
AC Q950A7:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DI 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cyan fluorescent protein (fragment).
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Scleractinia;
QC Pavlovina; Pavlovidae; Montastraea.
GX NCBI_TaxID:63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Falkowski P.G., Sun Y.
RT "Montastraea cavernosa fluorescent protein."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY056460; AAL17965.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01553; GFP: 1.
DR ProDom: PD013756; Green_fl_protein: 1.
FT NON_TER 225
FT TER 225
SQ SEQUENCE 225 AA: 25775 MW: 52DE2F716084524 CRC64.

```

```

Query Match 19.7%; Score 251.5; DB 5; Length 225;
Best Local Similarity 28.3%; Pred. No. 2.2e-13;
Matches 67; Conservative 48; Mismatches 81; Indels 41; Gaps 99

QY 11 VVPLVELDGVGHGKFSVSGEGEDADYVGRLEIRFICITLKLIVLWPTL 60
DB 12 VVPLVELDGVGHGKFSVSGEGEDADYVGRLEIRFICITLKLIVLWPTL 60
QY 70 CSRPYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
DB 71 CSRPYDHRKRDFFKSAMPEGVVYFRTIFFKCDQNYKTRASVREDELTLY 120
QY 67 VFTKVPKDIYDYFKQSPFSGYSWERSKTRGQVTVISIKLEIRGTFYETKRYGVN 124
DB 68 VFTKVPKDIYDYFKQSPFSGYSWERSKTRGQVTVISIKLEIRGTFYETKRYGVN 124
QY 130 FKEDGNILGHKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
DB 131 FKEDGNILGHKLFYNNYNSHNVYIMAKOKNSKVNFKIDNLEQNVLAQ 180
QY 125 FTSSGGVPMOKKTLKWPSTENYV-----ADGVLSGVNSFLIRGFT--KHHPNERS 175
DB 126 FTSSGGVPMOKKTLKWPSTENYV-----ADGVLSGVNSFLIRGFT--KHHPNERS 175
QY 189 GAGP-----VILPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
DB 190 GAGP-----VILPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
QY 176 TVGAKKGVLPVPEYFV-----GRIE-----LSHKKYNTVEVVE 311
DB 177 TVGAKKGVLPVPEYFV-----GRIE-----LSHKKYNTVEVVE 311

```

```

RESULT 15
Q876C0 PRELIMINARY: PRT: 236 AA.
AC Q876C0:
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DI 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
OS Dendronephthya sp. SSAL-2002.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Alcyonaria; Alcyonacea;
OC Nephtheidae; Dendronephthya.
GX NCBI_TaxID:191210;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEELINE-21927629; PubMed-11929995;
RA Lebas Y.A., Gurskaya N.G., Yanshevich Y.G., Fradkov A.F.,
RA Lukyanov K.A., Lukyanov S.A., Matz M.V.,
RT "Diversity and evolution of the green fluorescent protein family."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4256-4261(2002).
DR EMBL: AF420591; AAM10625.1;
SQ SEQUENCE 236 AA: 26840 MW: C61707FF934A90 CRC64.

Query Match 19.6%; Score 249.5; DB 5; Length 236;
Best Local Similarity 31.8%; Pred. No. 3.5e-13;
Matches 64; Conservative 39; Mismatches 85; Indels 33; Gaps 71

QY 14 TLVRIQDQVNGHGFYSVSGEGEDATVGRLEIRFICITLKLIVLWPTL 60
DB 15 VVPLVELDGVGHGKFSVSGEGEDADYVGRLEIRFICITLKLIVLWPTL 60
QY 75 KYEDHMKRHDFKSAFPEGVVQERTIFFKCDQNYKTRAEVKEGDTLVNKHGIDFKE 132
DB 76 KYEDHMKRHDFKSAFPEGVVQERTIFFKCDQNYKTRAEVKEGDTLVNKHGIDFKE 132
QY 70 EYFACDITDYFKQSPFSGYSWERTWYETKGTGTTIRSDTSLGDCGFQNIHFNKMFPP 127
DB 71 EYFACDITDYFKQSPFSGYSWERTWYETKGTGTTIRSDTSLGDCGFQNIHFNKMFPP 127
QY 133 DGNILGHKLFYNNYNSHNVYIMAKOKNGIKVNFKIRINIEDGGVGLAHVYGVN 190
DB 134 DGNILGHKLFYNNYNSHNVYIMAKOKNGIKVNFKIRINIEDGGVGLAHVYGVN 190
QY 126 NGPVMOKKTLKWPSTENYV-----ADGVLSGVNSFLIRGFT--KHHPNERS 175
DB 127 NGPVMOKKTLKWPSTENYV-----ADGVLSGVNSFLIRGFT--KHHPNERS 175
QY 191 GPVLPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
DB 192 GPVLPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
QY 181 KVGLPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240
DB 182 KVGLPDRHYLSTQSAQSKDNEKHEKVELEVFVAALIMHDELAK 240

```

Search completed: July 16, 2003, 16:13:29
 Job time : 41.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:13:38 : Search time 29 seconds
(without alignments)
974,649 Million CPU updates/sec

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGELFTGVVPIVLVDG.....VLIEFVIAAGI-HGMDELYK 238

Scoring table: RLOSUM62

Gapop 10.0 / Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 11899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications Aa*

1: /cgn2_6/prodata/2/pubaa/US07_NEW_PUB.pep.*
2: /cgn2_6/prodata/2/pubaa/PUB_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubaa/US06_PUB.pep.*
5: /cgn2_6/prodata/2/pubaa/US05_PUB.pep.*
6: /cgn2_6/prodata/2/pubaa/US07_PUB.pep.*
7: /cgn2_6/prodata/2/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubaa/US08_PUB.pep.*
9: /cgn2_6/prodata/2/pubaa/US09_NEW_PUB.pep.*
10: /cgn2_6/prodata/2/pubaa/US09_PUB.pep.*
11: /cgn2_6/prodata/2/pubaa/US10_NEW_PUB.pep.*
12: /cgn2_6/prodata/2/pubaa/US10_PUB.pep.*
13: /cgn2_6/prodata/2/pubaa/US6_NEW_PUB.pep.*
14: /cgn2_6/prodata/2/pubaa/US6_PUB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	DB ID	Description
1	1275	100.0	238	US-10-057-505-2
2	1275	100.0	238	US-09-884-561-2
3	1275	100.0	238	US-10-024-686-2
4	1271	99.7	238	US-09-866-538-2
5	1271	99.7	238	US-09-900-345A-12*
6	1271	99.7	238	US-10-121-358-16
7	1271	99.7	238	US-10-221-451-6
8	1271	99.7	238	US-09-920-922-4
9	1271	99.7	238	US-09-852-003-1
10	1268	99.5	243	US-09-900-345A-60
11	1268	99.5	243	US-09-900-345A-64
12	1268	99.5	243	US-09-900-345A-66
13	1268	99.5	243	US-09-900-345A-68
14	1268	99.5	243	US-09-900-345A-70
15	1268	99.5	1070	US-10-001-486B-2
16	1268	99.5	1099	US-10-259-864-4
17	1268	99.5	1147	US-10-259-864-7
18	1268	99.5	1147	US-10-259-864-7
19	1268	99.5	1452	US-10-050-673-2

20	1267	99.4	243	US-09-900-345A-54
21	1267	99.4	243	US-09-900-345A-56
22	1267	99.4	243	US-09-900-345A-58
23	1267	99.4	243	US-09-900-345A-114
24	1267	99.4	243	US-09-900-345A-116
25	1267	99.4	243	US-09-900-345A-118
26	1267	99.4	243	US-09-900-345A-120
27	1266	99.3	243	US-09-900-345A-2
28	1266	99.3	243	US-09-900-345A-4
29	1266	99.3	243	US-09-900-345A-6
30	1266	99.3	243	US-09-900-345A-8
31	1266	99.3	243	US-09-900-345A-10
32	1266	99.3	243	US-09-900-345A-12
33	1266	99.3	243	US-09-900-345A-14
34	1266	99.3	243	US-09-900-345A-16
35	1266	99.3	243	US-09-900-345A-18
36	1266	99.3	243	US-09-900-345A-20
37	1266	99.3	243	US-09-900-345A-22
38	1266	99.3	243	US-09-900-345A-24
39	1266	99.3	243	US-09-900-345A-26
40	1266	99.3	243	US-09-900-345A-28
41	1266	99.3	243	US-09-900-345A-30
42	1266	99.3	243	US-09-900-345A-32
43	1266	99.3	243	US-09-900-345A-34
44	1266	99.3	243	US-09-900-345A-36
45	1266	99.3	243	US-09-900-345A-38

ALIGNMENTS

RESULT 1
US-10-057-505-2
? Patent No. US062018467A1
? GENERAL INFORMATION:
? APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
? APPLICANT: AURORA BIOSCIENCES CORPORATION
? APPLICANT: TSIEH, Roger
? APPLICANT: HEIM, Roger
? APPLICANT: CURTIS, Andrew
? TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
? FILE REFERENCE: PEGN1260-3
? CURRENT APPLICATION NUMBER: US/10/057,505
? PRIOR FILING DATE: 2002-07-25
? PRIOR APPLICATION NUMBER: US 08/792,553
? PRIOR FILING DATE: 1997-01-31
? PRIOR APPLICATION NUMBER: US 09/596,003
? PRIOR FILING DATE: 1999-09-13
? NUMBER OF SEQ ID NOS: 31
? SOFTWARE: Patentin version 3.1
? SEQ ID NO: 2
? LENGTH: 239
? TYPE: PPT
? ORGANISM: Nequorea victoria
US-10-057-505-2

Query Match	100.0%	Score 1275	DP 9	Length 239
Best Local Similarity	100.0%	Pred. No. 1.5e-114		
Matches 239	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MSKGELFTGVVPIVLVDG	GVHGHKFSVSGEGDATY	GVKTLTKFICTGKLFVNPWTLL 60
DB	1	MSKGELFTGVVPIVLVDG	GVHGHKFSVSGEGDATY	GVKTLTKFICTGKLFVNPWTLL 60
QY	61	VTFYGVGVCFSRYPUHMKRHRD	FFKSAMPEGYVQERTIFFK	DGNYKTRAEVKFEGDTLV 120
DB	61	VTFYGVGVCFSRYPUHMKRHRD	FFKSAMPEGYVQERTIFFK	DGNYKTRAEVKFEGDTLV 120
QY	121	NRIELKGDIFKEDGKILGHKLE	YNTNSNYIMADKQNKVNP	KIRHNIELDSVQLAD 180
DB	121	NRIELKGDIFKEDGKILGHKLE	YNTNSNYIMADKQNKVNP	KIRHNIELDSVQLAD 180


```

GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSIEN, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530-2
CURRENT APPLICATION NUMBER: US/09/866,538
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-866-538-2

Query Match 99.7% Score 1271 DB 9: Length 238
Best Local Similarity 99.6% Pred. No. 3,5e-114
Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

CY 1 MSKGEELFTGVVPIVLVDGIVNGHKTSVSGEGEDATYKLTIKFKICTTKLTPKVPKPTL 60
DB 1 MSKGEELFTGVVPIVLVDGIVNGHKTSVSGEGEDATYKLTIKFKICTTKLTPKVPKPTL 60
CY 61 VTTFSGVQCFSRYPDHMKRHDFKSAKPEYGVQERTIFFKDGDNKYNKTRAEVKEEDTLV 120
DB 61 VTTFSGVQCFSRYPDHMKRHDFKSAKPEYGVQERTIFFKDGDNKYNKTRAEVKEEDTLV 120
CY 121 NRTELKGDKEKGNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNIEGSGVOLAD 180
DB 121 NRTELKGDKEKGNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNIEGSGVOLAD 180
CY 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238
DB 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238
CY 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238
DB 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238

RESULT 5
US-10-121-258-10
Sequence 10: Application US/0121258
Publication No. US20030059835A1
GENERAL INFORMATION:
APPLICANT: Tsién, Roger
APPLICANT: Campbell, Robert
TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
FILE REFERENCE: US083.13PZGPI
CURRENT APPLICATION NUMBER: US/10/21,258
CURRENT FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: 09/794,308
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 09/846,538
PRIOR FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 78
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-10-121-258-10

Query Match 99.7% Score 1271 DB 9: Length 238
Best Local Similarity 99.6% Pred. No. 3,5e-114
Matches 237: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

CY 1 MSKGEELFTGVVPIVLVDGIVNGHKTSVSGEGEDATYKLTIKFKICTTKLTPKVPKPTL 60
DB 1 MSKGEELFTGVVPIVLVDGIVNGHKTSVSGEGEDATYKLTIKFKICTTKLTPKVPKPTL 60
CY 61 VTTFSGVQCFSRYPDHMKRHDFKSAKPEYGVQERTIFFKDGDNKYNKTRAEVKEEDTLV 120
DB 61 VTTFSGVQCFSRYPDHMKRHDFKSAKPEYGVQERTIFFKDGDNKYNKTRAEVKEEDTLV 120
CY 121 NRTELKGDKEKGNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNIEGSGVOLAD 180
DB 121 NRTELKGDKEKGNILGHKLEYNNSHNYYIMADKKNKIKVNFKIRHNIEGSGVOLAD 180
CY 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238
DB 181 HYQONTPIGDSFVILPDNHYLSTQSALSKDPNEKPSHMVLEFVIAAGTTHGMDELYK 238

RESULT 7
US-10-221-451-6
Sequence 6: Application US/1022.461
Publication No. US20030092902A1
GENERAL INFORMATION:
APPLICANT: Marsh, Donald J.
TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR
TITLE OF INVENTION: CHIMERIC AND FUSION PROTEINS
FILE REFERENCE: 20652p
CURRENT APPLICATION NUMBER: US/10/221,461
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: PCT/US01/08071
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/189,698
PRIOR FILING DATE: 2000-03-15
NUMBER OF SEQ ID NOS: 37

```



```

Query Match      99.5%  Score 1268:  DB 9:  Length 243:
Best Local Similarity 99.2%:  Pred. No. 7c-114:
Matches 236:  Conservative 2:  Mismatches 0:  Indels 0:  Gaps 0:

QY 1 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 243

RESULT 11
US-09-900-345A-62
: Sequence 62, Application: US/09400345A
: Publication No. US20030031999A1
: GENERAL INFORMATION:
: APPLICANT: Frazer, Ian Hector
: APPLICANT: Zhou, Jian
: TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
: FILE REFERENCE: 10338-50S
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/09/900,345A
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 62
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)50FP
US-09-900-345A-62

Query Match      99.5%  Score 1268:  DB 9:  Length 243:
Best Local Similarity 99.2%:  Pred. No. 7c-114:
Matches 236:  Conservative 2:  Mismatches 0:  Indels 0:  Gaps 0:

QY 1 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 243

RESULT 12
US-09-900-345A-64
: Sequence 64, Application: US/09400345A
: Publication No. US20030031999A1

```

```

: GENERAL INFORMATION:
: APPLICANT: Frazer, Ian Hector
: APPLICANT: Zhou, Jian
: TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
: FILE REFERENCE: 10338-50S
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: US/09/900,345A
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 64
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)50FP
US-09-900-345A-64

Query Match      99.5%  Score 1269:  DB 9:  Length 243:
Best Local Similarity 99.2%:  Pred. No. 7c-114:
Matches 236:  Conservative 2:  Mismatches 0:  Indels 0:  Gaps 0:

QY 1 MSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 60
DB 6 LSKGEELFTGVVPIVLVDGDNVNGHKFSVSGSGGDDATYKLTLEFICTGKLPVFWPTL 65
QY 61 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTFSYGVOCFSRYPDHMKRHFDFKSAMPEGVVOERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 180
DB 126 NRLEKGIQDFKEDGNILGHKLEYNYNHSHVYIMADKQKNGKVNFKIRHNLEDGSSVQLAD 185
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 248
DB 186 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNKRDMHVLLEFVTAAGITHGMDELYK 243

RESULT 13
US-09-900-345A-66
: Sequence 66, Application: US/09400345A
: Publication No. US20030031999A1
: GENERAL INFORMATION:
: APPLICANT: Frazer, Ian Hector
: APPLICANT: Zhou, Jian
: TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
: FILE REFERENCE: 10338-50S
: CURRENT APPLICATION NUMBER: US/09/900,345A
: CURRENT FILING DATE: 2001-07-06
: PRIOR APPLICATION NUMBER: AU PP8078
: PRIOR FILING DATE: 1999-01-08
: PRIOR APPLICATION NUMBER: PCT/AU00/00008
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 185
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 66
: LENGTH: 243
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Leu(CTI)50FP
US-09-900-345A-66

Query Match      99.5%  Score 1268:  DB 9:  Length 243:
Best Local Similarity 99.2%:  Pred. No. 7c-114:
Matches 236:  Conservative 2:  Mismatches 0:  Indels 0:  Gaps 0:

```

QY 1 MSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 60
DB 2 LSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 65
QY 61 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 120
DB 66 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 125
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 180
DB 126 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 185
QY 181 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 240
DB 186 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 245

RESULT 14
US-09-900-345A-68
Query Match 99.5% Score 1268 DB 9 Length 243
Best Local Similarity 99.2% Pred. No. 70-114
Matches 236 Conservative 2 Mismatches 0 Indels 0 Gaps 0
GENERAL INFORMATION:
APPLICANT: Frazer, Ian Hector
TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
EFFICIENCY OF A CODON
FILE REFERENCE: 10338-508
CURRENT APPLICATION NUMBER: US-09/900-345A
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: AU PPR078
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCI/AJCC/00008
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 185
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 68
LENGTH: 243
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Leu(TTA)50FP
US-09-900-345A-68

QY 1 MSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 60
DB 2 LSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 65
QY 61 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 120
DB 66 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 125
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 180
DB 126 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 185
QY 181 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 240
DB 186 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 245

RESULT 15
US-09-900-345A-70
Query Match 99.5% Score 1268 DB 9 Length 243
Best Local Similarity 99.2% Pred. No. 70-114
Matches 236 Conservative 2 Mismatches 0 Indels 0 Gaps 0
GENERAL INFORMATION:
APPLICANT: Frazer, Ian Hector
TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
EFFICIENCY OF A CODON
FILE REFERENCE: 10338-508
CURRENT APPLICATION NUMBER: US-09/900-345A
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: AU PPR078
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCI/AJCC/00008
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 185
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 70
LENGTH: 243
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Leu(TTA)50FP
US-09-900-345A-70

FILE REFERENCE: 10338-508
CURRENT APPLICATION NUMBER: US-09/900-345A
CURRENT FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: AU PPR078
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PCI/AJCC/00008
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 185
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 70
LENGTH: 243
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Leu(TTG)53FP
US-09-900-345A-70

Query Match 99.5% Score 1268 DB 9 Length 243
Best Local Similarity 99.2% Pred. No. 70-114
Matches 236 Conservative 2 Mismatches 0 Indels 0 Gaps 0
QY 1 MSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 60
DB 2 LSKGEELTGVVPIVLVEIGDVGNGHKFSVSGEGDADATYGLKFKACTTOKLFPVWPTL 65
QY 61 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 120
DB 66 VTTFSYGVQCFSSRYPDHMKRHOFFKSAPEGYVOERTIFFKDGNGYKTRAEVKFEGDTLV 125
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 180
DB 126 NRLEKLGIDFKEDGNLGHKLEYNNSHNYYIMADKQKNGIKVNEKIRHNIEGSSVOLAD 185
QY 181 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 240
DB 186 HYQONTPIGDPVLLDHNHYLSTQSALSALSKDFNEKRDHMLLEFVTAAGTTHGMDELYK 245

Search completed: July 16, 2003, 16:30:10
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: July 16, 2003, 16:07:32 : Search time 20.5667 seconds

(without alignments)
1107.035 Million cell updates/sec

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELEFGVVPILVLELDS.....VILGFVTAAITTHKDEIVK 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 263224 seqs, 6618422 residues

Total number of hits satisfying chosen parameters: 263224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-731*

1: pir1*

2: pir2*

3: pir3*

4: pir4*

Fred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the best aligned protein,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	1265	99.2	238	1	C01514	green-fluorescent
2	1301	7.9	788	2	H72228	hypothetical prote
3	9515	7.5	1241	2	ER0048	coatomer complex e
4	95	7.5	336	2	C54468	hypothetical prote
5	91.5	7.2	887	2	E82590	luciferase-1
6	88.5	6.9	861	2	B54102	luciferase-1
7	87.5	6.9	862	2	AC0582	luciferase-1
8	87.5	6.9	2573	2	D71612	hypothetical prote
9	87	6.9	281	2	AD2052	hypothetical prote
10	87	6.9	357	2	G81455	hypothetical prote
11	87	6.8	471	2	D72856	hypothetical prote
12	87	6.8	632	2	T05586	RNA-binding protei
13	86	6.7	2334	2	S12920	cell wall-associat
14	85.5	6.7	836	1	J06211	DNA-directed DNA p
15	85.5	6.7	889	2	JC5376	inter-alpha-trypsi
16	85	6.7	647	2	H89988	hypothetical prote
17	84.5	6.6	370	2	E70350	iron-sulfur cofact
18	84.5	6.6	425	2	C97350	hypothetical prote
19	84.5	6.6	564	2	E81317	hypothetical prote
20	84.5	6.6	797	2	JC4078	protective surface
21	84.5	6.6	808	2	F64102	protective surface
22	84	6.6	529	2	B66815	hypothetical prote
23	84	6.6	760	2	F86281	protein F1086.14 i
24	83.5	6.5	461	2	T05936	hypothetical prote
25	83.5	6.5	616	2	C70931	hypothetical prote
26	83.5	6.5	1134	2	AC0234	tRNA for receptor pr
27	83.5	6.5	1164	1	FC0038	tRNA for receptor pr
28	83.5	6.5	2222	1	A36038	DNA-directed RNA p
29	83	6.5	207	2	C70483	hypothetical prote

RESULT: 1

C01514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 23-Mar-2001

C:Accession: J30692; J01534; PQ0335; S48693; S51330; S51331

K:Prasher, D.C.; Packeridge, V.K.; Ward, W.W.; Prasher, F.G.; Cormier, M.J.

Gene 111: 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: J01534; MUID:92175527; PMID:1347277

A:Accession: J30692

A:Molecule type: DNA

A:Residues: 1-107; S: 109-238 <PRA1>

A:Cross-references: GB:M62654; NCBI:555662; PDB:AAA27722.1; PDB:555663

A:Accession: C01514

A:Molecule type: mRNA

A:Residues: 1-99; F: 101-142; V: 142-218; V: 220-238 <PRA2>

A:Cross-references: GB:M62653; NCBI:555660; PDB:AAA27722.1; PDB:555661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-54; 74-122; 132-151; 154-183; 185-200 <PRA3>

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24; C: 25-156; P: 159-171; K: 173-238 <INP>

A:Cross-references: GB:L29345; NCBI:606183; PDB:AAA58246.1; PDB:606184

A:Watkins, J.N.; Campbell, A.K.

Submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Molecule type: mRNA

A:Residues: 1-13; V: 15-24; C: 26-44; N: 46-153; G: 155-156; P: 158-171; K: 173-227.

A:Cross-references: EMBL:883909; NCBI:634008; PDB:CAA58789.1; PDB:634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24; O: 26-29; R: 31-63; C: 65-103; G: 155-156; P: 158-171; K: 173-208.

A:Cross-references: EMBL:883960; NCBI:634010; PDB:CAA58790.1; PDB:634011

A:Experimental source: clone gfp2

K:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography; 1.9 angstroms, residues 'A' 2-79; 'R' 8

A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli

K:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

synaptotagmin 0.65
fructose-bisphosph
hypothetical prote
hypothetical prote
vegetative incomp
nitrogenase (EC 1.
oligoneuroneptidase
DNA topoisomerase
probable iron-cont
hypothetical prote
dihydrolypoamide d
alpha-amylase como
proteoglycan link
conserved hypothet
calpain 15C3.4.22
DNA-directed DNA p

A:Contents: annotation: X-ray crystallography, 2.5 angstroms
 C:Comment: This protein is excited by the photoreactive azobenzene (see F16-A21FNK) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-242 to Tyr-
 C:Genetics:

A:Gene: GFP
 A:Introns: 69/3: 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/Cross-link: 5-maleimide (Ser-242) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match: 99.2% Score 1265; DB 1; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1; Se-242;
 Matches 234; Conservative 4; Mismatches 0; Indels 0; Gaps 0

QY 1 MSKGBELFTGVVPILVELDPSVNGKPKSVSSGSRNENAIYFKITKEFITCKMLPQWNTL 60
 DB 1 MSKGBELFTGVVPILVELDPSVNGKPKSVSSGSRNENAIYFKITKEFITCKMLPQWNTL 60
 QY 61 VTIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 120
 DB 61 VTIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 120
 QY 62 VTIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 120
 DB 62 VTIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 120
 QY 121 NRLEKGLDSEKDNILGHLKLEYNNSHNYVMAQKQKNGIKVNFKIRHNEDGSVGLAD 180
 DB 121 NRLEKGLDSEKDNILGHLKLEYNNSHNYVMAQKQKNGIKVNFKIRHNEDGSVGLAD 180
 QY 181 HYQONTPIQDGGVLLPQNHLYLTSQALSQDKPKQKHVLLSEVTAANTTHGMDSLYK 240
 DB 181 HYQONTPIQDGGVLLPQNHLYLTSQALSQDKPKQKHVLLSEVTAANTTHGMDSLYK 240
 QY 191 HYQONTPIQDGGVLLPQNHLYLTSQALSQDKPKQKHVLLSEVTAANTTHGMDSLYK 240
 DB 191 HYQONTPIQDGGVLLPQNHLYLTSQALSQDKPKQKHVLLSEVTAANTTHGMDSLYK 240

RESULT 2
 H72228
 hypothetical protein TM1624 - Thermotoga maritima (strain MS8)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: H72228
 R:Neilsen, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.; Dodson, R.J.; Haft, D.H.; Hickox
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, G.A.; Richardson, D.
 C.M.

Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; PMID:99287456; PMID:11360671
 A:Accession: H72228

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1785 <XN>
 A:Cross-references: GR:AE001896; GR:AE000512; NCBI:94962196; PIR:AA046691.1; PIR:G498213
 A:Experimental: source: strain MS8
 C:Genetics:

A:Gene: TM1624

Query Match: 7.9% Score 101; DB 2; Length 745;
 Best Local Similarity 20.9%; Pred. No. 1.9;
 Matches 40; Conservative 27; Mismatches 64; Indels 6; Gaps 0

QY 2 SKGEELFTGVVPILVELDPSVNGKPKSVSSGSRNENAIYFKITKEFITCKMLPQWNTL 61
 DB 15 NEKRSFECTVPSVQAD 48
 QY 62 TIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 121
 DB 47 -----NEDKFEIDREWYERRFESKQWKSERVLVPSRWGLIS 88

QY 122 RIELKGLDSEKDNILGHLKLEYNNSHNYVMAQKQKNGIKVNFKIRHNEDGSVGLAD 181
 DB 89 DVLTLNGVY-----GSTEDMFIETPQVNTLV-----KEKNHLKVKYK-----STPVRK 134
 QY 182 YQNTPIQDGP 192
 DB 135 LEQNYGLVGP 145

RESULT 3

H72228

coarcter complex alpha chain homolog - human
 A:Alternate names: HEP-COP; xenopsin homolog; xenopsin-related peptide precursor
 N:Contains: xenin 25; xenopsin-related peptide
 C:Species: Homo sapiens (man)
 C:Date: 10-May-1996 #sequence_revision 06-Nov-1996 #text_change 21-Jul-2000
 C:Accession: J04668; A44317
 R:Kadow, V.T.K.; Quek, H.H.
 Gene 169, 223-227, 1996

A:Title: HEP-COP, a novel human gene whose product is highly homologous to the alpha-

A:Reference number: J04668; KJ10:96194906; PMID:8647451

A:Accession: J04668

A:Molecule type: mRNA

A:Residues: 1-1224 <XN>

A:Cross-references: GR:024105; NCBI:13877; PIR:AA070679.1; PIR:Q1002369

A:Experimental source: Hep4b hepatocellular carcinoma cell

R:Feurle, G.E.; Hamscher, G.; Kistek, R.; Meyer, H.E.; Metzger, J.W.

C: Biol. Chem. 267, 22305-22309, 1992

A:Title: Identification of xenin, a xenopsin-related peptide, in the human gastric mu

A:Reference number: A44317; KJ10:93054515; PMID:1429581

A:Accession: A44317

A:Molecule type: Protein

A:Residues: 1-25 <FE>

A:Experimental source: gastric mucosa

A:Note: sequence extracted from NCBI backbone (NCBI:117018)

A:Note: plasma levels of xenin 25, as determined by immunoassay, rise after meals and

A:Genetics:

A:Gene: GDB: COPA; HEP-COP

A:Cross-references: GDB:4642787; OMIM:601924

A:Map position: 1q23-1q25

A:Superfamily: yeast coarcter complex alpha chain; WD repeat homology

C:Keywords: duplication; hormone; plasma; stomach

F:1-25/Product: xenin 25 #status experimental <XN>

F:5-38/Domain: WD repeat homology <WD1>

F:17-25/Product: xenopsin-related peptide #status predicted <XRP>

F:47-80/Domain: WD repeat homology <WD2>

F:89-122/Domain: WD repeat homology <WD3>

F:131-164/Domain: WD repeat homology <WD4>

F:201-234/Domain: WD repeat homology <WD5>

F:245-278/Domain: WD repeat homology <WD6>

Query Match: 7.5% Score 95.5; DB 1; Length 1224;
 Best Local Similarity 24.6%; Pred. No. 9.7;
 Matches 48; Conservative 30; Mismatches 76; Indels 4; Gaps 0

QY 62 TIESGVGVCSRYPTTHMKRBUPEFKSAMPEGVGCKTIFFKQSGYKTAARVKSREGLV 122

DB 595 TEKFEKALLENKYEDEVLEKVRNAKLVQCSIIAYLQKKGYPEVALHFKYDE---KTFEFL 652

QY 111 KFFGDTLVNRIE LKGTDFEKEDGNILGHLKLEYNNSHNYVMAQKQKNGIKVNF----- 165

DB 652 ALKGNIEIAPAAKALDQKNEKLEVALQGNHOTIVKCYQRTKQKPKVSYLITG 712

QY 166 ---KIRENIEDSVQ---LADHYQONTPTGGGVLLPQNHLYLTSQALSQDKPKQKHVLL 220

DB 222 NLEKLRKNKIKARTRKQMSGHYQNALYGD-----VSEVRILKNGSKS----- 756

QY 221 LEFVTAAGTTHGMD 235

DB 757 LAYITAA--THGLGE 769

RESULT 4

C64466

hypothetical protein MJ1348 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: C64466
 R:Butt, C.; White, C.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Black
 Koech, C.J.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Marrick, J.M.; Gloeck,
 rson, J.D.; Sadow, P.W.; Hanta, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

```

Query Match      7.2% Score 91.5 DB 2 Length 887
Best Local Similarity 22.7% Pred. No. 14
Matches 45; Conservative 27; Mismatches 69; Indels 57; Gaps 10;

QY 49 TTGKLPVAPPTLVTFPSYGVQFSRYPDHMKRHDPEKFSAMPEGYVQERTIFFKDGNY-- 106
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 TNEQLPV-KVAVFVNAVGTGAVMVFGHDKQDEF--ANKYGLPIFOVIALKEFNODE 345
QY ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 -----KTRAEVKFEGDTLVNRTELKGDFFKEDGNTLGHKLKFNYSNHNVI 152
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 STWEPDVMWDYADKTR---EFE--LINSAPEDGLDYQDAPEVLAERFE ----- 429
QY 153 MADKQKNG-IKVNFKTRHNIEDGSVOLADHYQNTT-----GCSFVLLPON 198
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 ---RQGGORRYNRLR---DMGVSQRQYWGCVPIVIVYCTCCGAVVPEDQLVLPEN 482
QY 199 -HYLSQTQSALSKDPNEKR 215
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 VAFSTGTGSPIKTDPERKK 500

RESULT 6
H64102
leucyl-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Ki KW29);
N Alternate names: leucyl-tRNA synthetase
C Species: Haemophilus influenzae
C Date: 18-Aug-1995 #sequence, revision 18-Aug-1995 #text, change 03-Jun-2002
C Accession: H64102
R Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, P.A.; Kirkness, E.F.; Kerl,
  G.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.; Godek, A.; Kelley, J.M.; W.
  D.M.; Brandon, R.C.; Fink, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen,
  Science 269, 496-512, 1995
A Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
A Reference number: A6400C; KUID:95350630; PMID:7542800
A Accession: H64102
A Start: nucleic acid sequence not shown; translation not shown
A Molecule type: DNA
A Residues: 1-261 <PIGA>
A Cross-references: GR:032774; GR:542024; NID:gl574942; PID:AA22951.1; PID:gl2
C Genetics:
C Superfamily: leucine-tRNA ligase
C Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match      6.9% Score 88.5 DB 2 Length 861
Best Local Similarity 24.3% Pred. No. 24
Matches 45; Conservative 25; Mismatches 68; Indels 47; Gaps 9;

QY 49 TTGKLPVAPPTLVTFPSYGVQFSRYPDHMKRHDPEKFSAMPEGYVQERTIFFKDGNYKI 108
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TGDKLPV-KVAVFVNAVGTGAVMVFGHDKQDEF--PEFAQKSLPIKQVTAPEIDQL 370
QY 109 EAEVKFEGDTLVNRTELKGDFFKEDGNTLGHKLKFNYSNHNVIADK QKNEI---KVN 164
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 TQGAPEVHEKLVNSDEEDGKNE--DGAENS-----IADKLEKLVGKRGQVN 414
QY 165 FKLRH-----NTEDSVQASHYQNTPTGDFVLLPQNHYL-STUSAL 207
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 YRLRDGVSQRQYWGCVPIVIVYCTCCGAVVPEDQLVLPEN 467
QY 208 SKDPN 212
DB ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
468 NADPN 472

RESULT 7
AC0582
leucyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar
C Species: Salmonella enterica subsp. enterica serovar typhi
A Note: this species has also been called Salmonella typhi
C Date: 09-Nov-2001 #sequence, revision 09-Nov-2001 #text, change 27-Nov-2001

```

C;Accession: AC0582

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.A.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.K.; Dowd, L.; White, N.; Farrar, S.; Mout, S.; G'Gaora, P.
 Nature 413, 848-852, 2001
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
 A:Reference numbers: AB0502; PMID:11677608
 A:Accession: AC5582
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-860 <PAR>
 A:Cross-references: GR:AL513362; PIRN:CA045126.1; PIRN:Q1601349; GSTER:GN00179
 A:Genetic: STV0899
 A:Superfamily: leucine-tRNA ligase
 C:Query Match 6.0%; Score 67.5; DB 2; Length 860;
 Best Local Similarity 22.0%; Pred. No. 20;
 Matches 41; Conservative 21; Mismatches 79; Indels 39; Gaps 6;
 QY 49 TTGKLVFNTLV:TSYGVGVPSFVPEHKKPHPEKKSAMPGYVQSTIFPKLQNKI 10P
 DB 314 TGEIVV-KAANFVCMYGTGAMVATGDCG--YEFASKVGLTIKIVLAAGSEIL 40
 QY 109 RAEVKEEMTLVNR:ELKQIIEKEDAGNLSHMKLENNYNSNVYMAKQKQKYNKIK 168
 DB 371 SGAETKGVVNSJPFDDGATGAHNAIAPL-----AKVGVSEKYNKIF 418
 QY 149 H-----NTGASVGLADHYQNTFPGHGVLPNNHYL-STGSAISKIP 211
 DB 419 DMGVSQRYGAPIPAVTIFEDGTV-----LPTFEDQLV:LPENVMVGSTPIKASP 471
 RESULT 9
 D71614
 Hypothetical protein PF04500 - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 A:Accession: D71614
 R:Gardner, M.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 Pierce, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.G.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference numbers: A71600; M01D:99021743; PMID:9804551
 A:Accession: D71614
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2573 <PAR>
 A:Cross-references: GB:AE001376; GR:AE001362; NID:04545186; PIRN:AA071581.1; PIR:Q184519
 A:Experimental source: clone 307
 A:Genetics:
 A:Gene: PF04500

Query Match 6.0%; Score 57.5; DB 2; Length 2573;
 Best Local Similarity 26.2%; Pred. No. 1,226-02;
 Matches 34; Conservative 30; Mismatches 53; Indels 13; Gaps 5;
 QY 93 VQRTVFPRD--CGNYKTRAEVFRDGLVNR:KGLDFKNGNLGKLEVN--YKSH 148
 DB 126 LKRETLKRT:KSSNDPMD:SLFDQDVQKELK--DFEKSLSKTKKKEVNTFLYKKA 183
 QY 149 NYTIMACKQKNGTKVKNKTRHN:EDGSVGLADHYQNTFPGHGVLPNNHYL-STGSAISKIP 208
 DB 184 NLHKENKKKQKQKKNK:HNNDNNM----LYVKNL---(KTRV)LENNVNVHTLNCN 236
 QY 209 KDPNKECHM 216
 DB 237 TYLKRERDYM 246

RESULT 9
 AD2052
 Hypothetical protein all1970 (imported) - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AD2052
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuri, T.; Sasagawa, S.; Watanabe, A.; Iizumi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference numbers: AB1907; M01D:21595285; PMID:11759640
 A:Accession: AD2052
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-281 <KOR>
 A:Cross-references: GB:BA000019; PIRN:BA073669.1; PIR:Q1731060; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 A:Genetics:
 A:Gene: all1970

Query Match 6.0%; Score 97; DB 2; Length 281;
 Best Local Similarity 24.4%; Pred. No. 7;
 Matches 33; Conservative 26; Mismatches 54; Indels 22; Gaps 6;
 QY 106 YKTRAEVFRDGLVNR:KGLDFKNGNLGKLEVN--YKSH 148
 DB 126 LKRETLKRT:KSSNDPMD:SLFDQDVQKELK--DFEKSLSKTKKKEVNTFLYKKA 183
 QY 149 NYTIMACKQKNGTKVKNKTRHN:EDGSVGLADHYQNTFPGHGVLPNNHYL-STGSAISKIP 208
 DB 184 NLHKENKKKQKQKKNK:HNNDNNM----LYVKNL---(KTRV)LENNVNVHTLNCN 236
 QY 209 KDPNKECHM 216
 DB 237 TYLKRERDYM 246

RESULT 10
 GB1355
 tRNA (tyracyl-5)-methyltransferase (EC 2.1.1.35) Cj0831c (imported) - Campylobacter j
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 A:Accession: GB1355
 R:Parkhill, J.; Wren, B.W.; Mungai, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Hill,
 C.W.; Ouali, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference numbers: AB1250; M01D:2015912; PMID:10588204
 A:Accession: GB1355
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-357 <PAR>
 A:Cross-references: GB:AL139076; GR:AL111168; NID:06968128; PIRN:CA573090.1; PIR:Q496
 A:Experimental source: serotype O2, strain NCTC 11166
 A:Genetics:
 A:Gene: tRNA; Cj0831c
 C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.8%; Score 87; DB 2; Length 357;
 Best Local Similarity 24.8%; Pred. No. 10;
 Matches 30; Conservative 18; Mismatches 39; Indels 34; Gaps 5;
 QY 79 KRHOFFKAMPEGVQVQRT:FFKDGNGYKTRAEVFRDGLVNR:KGLDFKNGNLGKLEVN--YKSH 148
 DB 126 LKRETLKRT:KSSNDPMD:SLFDQDVQKELK--DFEKSLSKTKKKEVNTFLYKKA 183
 QY 149 NYTIMACKQKNGTKVKNKTRHN:EDGSVGLADHYQNTFPGHGVLPNNHYL-STGSAISKIP 208
 DB 184 NLHKENKKKQKQKKNK:HNNDNNM----LYVKNL---(KTRV)LENNVNVHTLNCN 236
 QY 209 KDPNKECHM 216
 DB 237 TYLKRERDYM 246

RESULT 9
 AD2052
 Hypothetical protein all1970 (imported) - Nostoc sp. (strain PCC 7120)

GenCode version 5.1.4
Copyright (c) 1993 - 2003 Comptech Ltd.

OM protein - protein search, using sw mode:

Run on: July 16, 2003, 10:01:23 : Search time 12.6667 seconds

(without alignments)
925.441 Million cell updates/sec

Title: US-09-967-X01-2

Perfect score: 1275

Sequence: 1 MSKGELEFGVVPILDELGG.....VLLEFVTAG:TRXMDLYK 238

Scoring table: BLOSUM62

Gapop 10.0 : Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match OK

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT_40.*

Pred. No. is the number of residues predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	OM ID	Description
1	1271	99.7	238	GFP_AEQVI	
2	95.5	7.5	1224	COXA_HUMAN	
3	95	7.5	936	YDAR_MITJA	
4	92.5	7.3	1224	COXA_BOVIN	
5	91.5	7.2	879	SVL_XLEPA	
6	88.5	6.9	861	SVL_HABIN	
7	87	6.6	837	TRMA_CAME	
8	86	6.7	2734	WAPA_HACSC	
9	85.5	6.7	533	CP51_CAMEA	
10	85.5	6.7	486	ITHI_MPSA7	
11	84.5	6.6	795	G162_HABIN	
12	84.5	6.6	797	D151_HABIN	
13	83.5	6.5	441	PSB2_COAPA	
14	83.5	6.5	733	D152_HABIN	
15	83.5	6.5	1164	BAG_STRAG	
16	83.5	6.5	2222	DPOE_YEAST	
17	83	6.5	439	SY62_DFSOM	
18	82.5	6.5	533	N1P0_GLOPA	
19	82.5	6.5	513	PEPF_MYCP9	
20	82	6.4	774	AMV2_SCHPO	
21	81.5	6.4	355	PLK_CHECK	
22	81.5	6.4	333	MTG4_HCLPV	
23	81.5	6.4	700	CAN2_HUMAN	
24	81.5	6.4	788	SPO2_HPBHE	
25	81	6.4	836	DF01_RPB09	
26	81	6.4	953	YNM7_YEAST	
27	80.5	6.3	324	CAN2_PIG	
28	80.5	6.3	752	NEC1_RAT	
29	80.5	6.3	860	SVL_ECOLI	
30	80	6.3	297	CAN2_BOVIN	
31	79.5	6.2	217	EXPI_ERWCA	
32	79.5	6.2	422	CAN2_RABIT	
33	79.5	6.2	458	GLNA_AZOCA	

RESULT 1

GFP_AEQVI

ID GFP_AEQVI STANDARD: PRT: 238 AA.

AC P42212: Q37104:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 14-OCT-2001 (Rel. 40, Last annotation update)

DE Green fluorescent protein.

GN GFP.

OS Aequorea victoria (Jellyfish).

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoriidae; Aequorea.

OX NCBI_TaxID=6100;

RN [1]

RF SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

FX MEDLINE=92175527; PubMed=1347377;

RA Prasher D.C., Eckhardt V.K., Ward W.W., Prendergast P.G.,

RA Corbiter M.S.;

RT "Primary structure of the Aequorea victoria green-fluorescent

RT protein."

RL Gene 11:229-233(1992).

RN [2]

RF SEQUENCE FROM N.A.

FX MEDLINE=94195810; PubMed=8187953;

RA Inouye S., Tsuji F.;

RT "Aequorea green fluorescent protein. Expression of the gene and

RT fluorescence characteristics of the recombinant protein."

RL FEBS Lett. 341:277-280(1994).

RN [3]

RF CHROMOSOMES.

FX MEDLINE=93192221; PubMed=8442132;

RA Cody C.W., Prasher D.C., Webster W.M., Prendergast P.G., Ward W.W.;

RT "Chemical structure of the hexapeptide chromophore of the Aequorea

RT green-fluorescent protein."

RL Biochemistry 32:1212-1218(1993).

RN [4]

RF X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

FX MEDLINE=96355655; PubMed=8703075;

RA Ormrod M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.;

RT "Crystal structure of the Aequorea victoria green fluorescent

RT protein."

RL Science 273:1392-1395(1996).

RN [5]

RF X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

FX MEDLINE=98294543; PubMed=9631087;

RA Yang F., Moss L.G., Phillips G.N. Jr.;

RT "The molecular structure of green fluorescent protein."

RL Nat. Biotechnol. 14:1246-1251(1996).

RN [6]

RF X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.

FX MEDLINE=98455509; PubMed=9782051;

RA Wachter R.M., Elsflinger M.A., Kallio K., Hanson G.T., Remington S.J.;

RT "Structural basis of spectral shifts in the yellow-emission variants

RT of green fluorescent protein."

RL Structure 6:1267-1277(1998).

Q63416 ratfus mrv
Q48449 klebsiella
P49052 bacillus b
P22005 mus muscula
Q88500 rattus norv
Q00116 homo sapien
P39321 escherichia
Q94101 sus scrofa
Q94107 chlamydia m
P42656 spongilla
P36924 bacillus co
Q08529 mus muscul

ALIGNMENTS

[2]
 RP SEQUENCE OF 31-50: 446-472 AND 504-523, AND SUBUNIT.
 RC TISSUE-Plasma:
 RA MEDLINE-97018241; PubMed-8854857;
 RT "inter-alpha-trypsin inhibitor and its related proteins in Syrian
 RL C. Biochem. 120:145-152(1996).
 CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SEMIN OR AS A
 CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,
 CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
 CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
 CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: 1-ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM
 CC ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN.
 CC BIKUNIN, INTER-ALPHA-INHIBITOR (1-ALPHA-1) IS COMPOSED OF H1, H2
 CC AND BIKUNIN, INTER-ALPHA-LINE INHIBITOR (1-ALPHA-2) OF H2 AND
 CC BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-1) OF H3 AND BIKUNIN.
 CC -1- PTM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CROSSLINK
 CC 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARATE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TIRI FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 WFA DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database,
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

EMBL: D89287; AAC1340.1;
 DR InterPro: IPR002035; WFA;
 DR Pfam: PF00092; wfa; 1;
 DR SMART: SM00327; WFA; 1;
 DR PROSITE: PS00234; WFA; 1;
 KW Serine protease inhibitor; Repeat; Signal: Multidomain family;
 KW Glycoprotein;
 FT SIGNAL: 1 16 POTENTIAL;
 FT PROPEP 19 30 BY SIMILARITY;
 FT CHAIN 31 646 IN-TER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
 H3;
 FT PROPEP 647 866 BY SIMILARITY;
 FT DOMAIN 279 439 WFA;
 FT CARBOHYD 88 88 N-LINKED (GLCNAC...) (POTENTIAL);
 FT CARBOHYD 577 577 N-LINKED (GLCNAC...) (POTENTIAL);
 FT BINDING 545 646 CHONDROITIN 4-SULFATE, CROSS-LINK SITE
 (BY SIMILARITY).
 SC SEQUENCE 886 AA; 99018 MW; AC0594C6852576B CRC64;
 Query Match 6.78; Score 85.9; DB 1; Length 886;
 Best Local Similarity 23.84; Pred. No. 14;
 Matches 36; Conservative 34; Mismatches 52; Indels 19; Gaps 7;
 QY 74 YFHKRKHDFKSAPEGVVQERT-----TEFKDGNGVKRAEVKPEGD---TLVNRIE 124
 DB YPENALILDLTKNSYPHEYGSEFAVAGKADSLMNNK--NPKVHGALNLTFTFEVD 512
 QY 135 LKGIQD--FKESGNIQKLENNYNSHRVYINAKOKNKLKYNKTKRHNLEYSVGLADHY 142
 DB 543 KKEHMAALKEGYIFRNYIEHMAVITTEQLLEKRNK--KARDEKFNELACLESSEHY 564
 QY 163 QNTPHIDGPPVLLTUNYLSQSALSNDNE 213
 DB 590 HFFVPLTPTKVTKEDN--EDPTSLACKYGE 616
 RESULT 12
 ID D152_HAFIN STANDARD; PRI: 795 AA;
 AC P44935;

Q1-NOV-1995 (Rel. 32, Created)
 Q1-NOV-1995 (Rel. 32, Last sequence update)
 D1 15-JUN-2002 (Rel. 41, last annotation update)
 DB Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (p15-Aq) (Outer membrane protein D15).
 GN H19917.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma sub-division; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1] TaxID=727;
 SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RA MEDLINE-8535033; PubMed-7542800;
 RA Fleischman R.D., Adams M.D., White O., Clayton R.A., Kirkness P.F.,
 RA Kerlavage A.R., Bitt C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu J.-L., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nuygen D.T., Saudek D.M., Brandon N.G.,
 RA Fine L.D., Fritchman J.B., Fritchman J.L., Geoghegan N.S.M.,
 RA Gnehm C.E., McDonald L.A., Small K.V., Fraser C.M., Smith H.J.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RI Rd.";
 RI Solene 89:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database,
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

EMBL: U32773; AAC22575.1;
 DR TISH: H20517;
 DR InterPro: IPR000184; Bac_surface_Ag_1;
 DR Pfam: PF01103; Bac_surface_Ag; 1;
 KW Antigen; Outer membrane; Signal; Complete proteome.
 FT SIGNAL: 1 19 POTENTIAL;
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
 SC SEQUENCE 795 AA; 87478 MW; B85691FC225E6D4 CRC64;
 Query Match 6.58; Score 84.5; DB 1; Length 795;
 Best Local Similarity 21.78; Pred. No. 14;
 Matches 47; Conservative 31; Mismatches 76; Indels 63; Gaps 12;
 QY 66 YGVCFGRYPDHRKHDF-----PKSAMPEGVVQ-----PIIFF 100
 DB 429 YGTSGTSYQASVKODNPLGTGAASVLAAGTGNKGYISVNLGYTEPYFTKGVSLGNVFP 488
 QY 101 KDDGNKTRAEVKPEGDTLVNRIELKGLDKEDGNI---LGH-----KLEYNYSIN 149
 DE 489 EYNDKSKDTSNKRTTYGTSNVTL-GPPVNNKNSYVVGHTYNTKISAEALYN---PN 544
 QY 150 VYIMADKQK-NGLKVNKIRHNJEDGSGVGLADHYQ-----NTPKSSVLLP- 196
 DB 545 EYISMKFKGNGHILTN-----DEDFSGWNNISLNGYFTTKGVKASLT-GRVING 595
 QY 197 -DNHYLSQSALSNDNEKPKNEKHMVLLFEVTAAGITHS 232
 DE 596 SDNYYKLSADVGQGYFDLDRHLWVWSAKASAGYANG 632
 RESULT 12
 ID D151_HAFIN STANDARD; PRI: 797 AA;
 AC P46024;
 Q1-NOV-1995 (Rel. 42, Created)
 Q1-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DN D15-Ag; Outer membrane protein D15;
CS Haemophilus influenzae;
CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
CC Haemophilus;
CX NCBI_TaxID=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype B;
RC MEDLINE=95253676; PubMed=7737523;
RA Flack F.S., Loomore S., Chong P., Thomas W.R.;
RT "The sequencing of the 60-kDa D15 protective surface antigen of
RT Haemophilus influenzae.";
RI Gene 156:97-99(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Minna, and Eagan / Serotype B;
RX MEDLINE=97427952; PubMed=728140;
RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.K., England L.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RI Infect Immun 65:3701-3707(1997).
CC (1) SUBCELLULAR LOCATION: Outer membrane.
CC (2) SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch.)
DR EMBL: U13961; AA85845.1;
DR EMBL: U60832; AB61974.1;
DR EMBL: U60833; AB61976.1;
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Flun: PF01103; Bac_surfAg_Ag_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA: 67675 MW: 259305.466 kDa: B 67675;
Query Match 6.68; Score 84.5; DP 1; Length 797;
Best Local Similarity 20.73; Pred. No. 15;
Matches 47; Conservative 33; Mismatches 76; Indels 64; Gaps 11;
QY 66 KVVQFNSYFPHMKRHUP----- --FKSAMPSEYWE-----ELIF 170
DY 429 KATSGISQASQKDNFLGGAASVATKNDAGISVAVYEDAFYKDVSLKVEF 486
QY 101 KKGNGKRAEVKFGDGMRIKRLKEDKDNV----- --KLFYKNSN 544
DY 489 ENYDNRSDTSNKKATYGSNVLTGFFNNKSTYVGLDHTYKNSNFAEYN-- -- 544
QY 130 VYIMADKOK-NEIKVFAIRNIEDSGVOLADHYQV----- --NTPTGQVYLD- 596
DY 545 LYIQMKFKGNGIKTN----- --DFPSTGKNNLSLKKVPEYCKGVKASLG-GRYTI 545
QY 197 -DNHYLSTQSALSKDPNKKRDMYLLFTVFNGLTHG 242
DY 596 SDNKKYKLSADVQGYPLDROHJWVSKAKASGYANG 432
RESULT 13
PSBC_CVAPA STANDARD; PRT: 46; AA:
AC P48104;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
DN PSRC.
CS Cyanophora paradoxa.
CC Cyanelle.
CC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
CX NCBI_TaxID=3762;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=L5555 / Pringsheim;
RA Stirewalt V.L., Michalowski C.B., Ioffelhardt W., Bohrer H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RI Plant Mol. Biol. Rep. 13:327-332(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=L5555 / Pringsheim;
RA Ioffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakovitsch J., Bohrer H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RI (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RI Schwemler W. (eds.);
RI Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RI (1997).
CC (1) FUNCTION: THE 43 KDA PROTEIN (P6) IS A COMPONENT OF THE CORE OF
CC PHOTOSYSTEM II. IT IS A CHLOROPHYLL BINDING PROTEIN.
CC (2) SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CYANELLE
CC THYLAKOID MEMBRANE.
CC (3) SIMILARITY: BELONGS TO THE PSBB / PSRC FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch.)
DR EMBL: U30821; AA81279.1;
DR InterPro: IPR000932; PSIIProt.
DR Flun: PF00421; PSI1.1;
DR TrGFAMS: TIGR01153; PSBC_1.
KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Cyanelle;
KW Transmembrane.
SQ SEQUENCE 451 AA: 50402 MW: 42664E249E3B9C1 CRC64;
Query Match 6.58; Score 83.5; DP 1; Length 451;
Best Local Similarity 20.24; Pred. No. 9.1;
Matches 52; Conservative 33; Mismatches 90; Indels 83; Gaps 10;
QY 24 GHRFSVSGEGEATYKILTKLICTGTGKIPVWPTLVTFSYG----- -- 67
DY 206 GUGNIVSVQKMDIISGHVLAFCILGCV---WHITKPSKARALWNSSEAVLSYL 264
QY 68 -----VQCSRPYDDHMKRHDFTKSNAPG-GYVQERTIFFKDD----- 103
DY 225 AALALMGFIANCFYWFNNIAYPSEFFOPTSPASQAFITFLVFEORCANVWSAGPTG 424
QY 104 -GNVKTAA---EYKFEEDTIVNNIELKGIDFKDGNILGHKLEYNVNSHVYIMADKKN 159
DY 325 LGKYLKMRSPSGEITIFGETM----- --KWDTRAPKLEPGRAN 361
QY 160 GKVNFKIRHNIEDSGVOLADHYQVNTPIGD----- --GPVLSPDN--HYLSTQSALSKDNE 213
DY 362 GDLT-KIKYDIQFQWERRAAEYNIHAPLSLNSGVGVAETNSVNSVSPREWS----- 415
QY 214 KPDHMYLLEFVTRAGTH 231
DY 416 -TSHFVLGFLFIGHLWH 432

RESULT 14

D153_HAETN STANDARD: PRT: 793 AA.
AC Q32629;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa pI; antigen)
DE (3-15-Aq) Outer membrane protein D15
OS Haemophilus influenzae
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales
OC Haemophilus
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PAK 12085;
RX MEDLINE=9742792; PubMed 9284140;
RA Locmore S.M., Yang Y.F., Coleman D.C., Shortreed J.M., Norland D.M., Klein M.H.
RT Outer membrane protein D15 is conserved among Haemophilus influenzae species and may represent a universal protective antigen against invasive disease.
RL Infect. Immun. 65:3731-3737(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).
CC -----
CC EMBL: U65834; AAB61977.1;
CC InterPro: IPR00184; Bac_surfAq_D15.
CC Pfam: PF01103; Bac_surface_Aq1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA: 87511 MW: 51862036601A14 CRC64:

Query Match 6.5%: Score 83.5; DB 1; Length 793;
Best Local Similarity 22.1%: Pred. No. 18;
Matches 48; Conservative 20; Mismatches 77; Indels 83; Gaps 11;
QY 66 YGVQCFSRVPEHMKRDF-----FKSAMPEGVQE-----KTIFF 160
DE 429 YGTESG:SVYTS:KQNFELGTGAAVTAG:KNYH:SVNLGY:EPYFK:GVSLGNIFF 498
QY 101 KQGNKYTRAEVFEEDTLVNRLELKGIDPK:KGN:---LGH-----KLEYNYSNR 149
DE 489 ENYDNRKSTSSNYKRTYGSNVL:GFVYNNNSVYVUGHEVYKTSNFALRYN:--RN 544
QY 150 VYIMAKQK:NGIKVNFKISHNEDGNSVOLADYQQ:---NTIIGGPEVLLP: 156
DE 545 LYQSMKFKNGKNTN-----DEFSQKYNLSNRPYETKVKVAKSLR:GEVTIFG 545
QY 197 -DNRYLSTQALSQRNEKTHVLEFVIAATGES 642
DE 596 SNKYYKLSADVGQGYPLDCHRWVSKAKASAVAK 642

RESULT 15

BAG_STRAG
ID BAG_STRAG STANDARD: PRT: 1164 AA.
AC Q27951;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Iga FC receptor precursor (Beta antigen) (B antigen).
GN BAG.
OS Streptococcus agalactiae.

CC Bacteria; Firmicutes; Actinobacillales; Streptococcaceae;
CC Streptococcus;
CX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-43.
RX STRAIN=LA399;
RX MEDLINE=91312121; PubMed 1857207;
RA Jurlstrom P.G., Chhatwa G.S., Timmis K.N.;
RT "The Iga-binding beta antigen of the C protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions".
RI Microbiol. 9:843-849(1991).
RN [2]
RP IDENTIFICATION OF IG-LIKE DOMAIN.
RX MEDLINE=9733265; PubMed=8880921;
RA Bateman A., Paddy S.R., Chothia G.;
RT "Members of the immunoglobulin superfamily in bacteria".
RI Protein Sci. 5:1439-1442(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).
CC -----
CC EMBL: X53771; CAA42442.1;
CC PIR: S5330; FCSGAG.
CC InterPro: IPR04829; Csurface_antigen.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR003599; Ig.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC ProDom: PD153432; Csurface_antigen; 2.
CC SMART: SMO0409; Ig; 1.
CC TIGRFAMs: TIGR01167; LPXIG_anchor; 1.
CC TIGRFAMs: TIGR01168; YSIRK_signal; 1.
CC PROSITE: PS00847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal;
KW Immunoglobulin domain.
FT SIGNAL 1 37
FT CHAIN 38 1135 IGA FC RECEPTOR.
FT PROPEP 1136 1164 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 434 534 IG-LIKE DOMAIN.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT DOMAIN 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT SITE 1132 1136 LPXIG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1135 1135 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1164 AA: 131051 MW: 650594AF720A5474 CRC54:

Query Match 6.5%: Score 83.5; DB 1; Length 1164;
Best Local Similarity 24.5%: Pred. No. 28;
Matches 39; Conservative 24; Mismatches 65; Indels 81; Gaps 7;
QY 73 RYPCMKRHRDFKSAPEGVQ:ERTIFFKDCNYKTRAEVFEEDHIVNA:TEIKGIDPK 131
DE 930 KLPDVPKLPDAPK--LPQINKVGAVPTSDGN--TKVTTFDQPTDAKLLHLEVTK 985
QY 132 EDGNLIGHKJ-----EYNNSHNVYIMADKQNGIKVNFKIRNIEDG 174
DE 996 ELADKIAHKTGSGTVKRVFDSLKSGKETHVNGERTVRLALQGTGS---DVHYVVKENG 1042
QY 175 SVOLADHYQONTPLSDGFPVLLPDNIH--LSTQSALSKDPN 212
DE 1043 DLERIPSKVEN-----QGVVFKTNHFSFAIKTLKSDQN 1076

Search completed: July 16, 2003, 16:11:17
Job time : 13.6667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer 201

us protein - protein search, using sw model

Run on: July 16, 2003, 16:06:07 ; Search time 41.334 seconds
(without alignment)
1186.43. Billion cell updates/sec

Title: US-09-967-301-2
Perfect score: 1275
Sequence: 1 MSKGEELFGVVPILVELDQ.....VILEPVCAGITHEDELYK 238

Scoring table: BLOSUM62

Gapop 15.0 , Gapext 3.5

Searched: 671560 seqs, 266047115 residues

Total number of hits satisfying chosen parameters: 671590

Minimum DB seq length: 6

Maximum DB seq length: 200660000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_2.1.1
1: sp_archaea
2: sp_bacteria
3: sp_fungi
4: sp_humani
5: sp_invertebrate
6: sp_mammali
7: sp_mol
8: sp_mol
9: sp_mol
10: sp_mol
11: sp_mol
12: sp_mol
13: sp_mol
14: sp_mol
15: sp_mol
16: sp_mol
17: sp_mol

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	Description
1	1271	99.7	238	5	Q27903: unclassified
2	1264	99.1	238	5	Q3125: aequorea vi
3	1229	96.4	238	5	Q17105: aequorea vi
4	1214	95.2	238	5	Q17106: aequorea vi
5	1211	95.0	758	4	Q96701: aequorea vi
6	1108	86.9	238	5	Q8WTC6: aequorea ma
7	1102	86.4	238	5	Q8WTC6: aequorea ma
8	1094	85.8	238	5	Q8WTC6: aequorea ma
9	1091	85.6	238	5	Q8WTC6: aequorea ma
10	1091	85.6	238	5	Q8WTC6: aequorea ma
11	1091	85.6	238	5	Q8WTC6: aequorea ma
12	1089	85.4	238	5	Q8WTC6: aequorea ma
13	1087	85.3	238	5	Q8WTC6: aequorea ma
14	259.5	20.4	225	5	Q950A7: montastraea
15	253	19.8	225	5	Q963F5: montastraea
16	244.5	19.2	266	5	Q906Y3: clausuaria

17	243.5	18.9	236	5	Q8T5C0: condronapat
18	238	18.7	225	5	Q8T5F1: montastraea
19	218	16.9	229	5	Q905Y6: anemonia ma
20	214.5	16.8	234	5	Q8T5F2: montastraea
21	213.5	16.7	235	5	Q8T5F0: scyllamia cu
22	212	16.6	227	5	Q8T5P9: montastraea
23	212.5	16.5	225	5	Q905Y8: discosoma s
24	205.5	16.4	238	5	Q905Y9: montastraea
25	208	16.3	227	5	Q905Y0: montastraea
26	205	16.1	232	5	Q905Y7: discosoma s
27	204.5	16.0	232	5	Q905Y5: anemonia su
28	204	16.0	221	5	Q905Y4: anemonia su
29	203.5	16.0	225	5	Q8T5G9: heteractis
30	205.5	15.7	232	5	Q8G228: anemonia su
31	196	15.4	230	5	Q905Y7: discosoma s
32	183.5	15.2	228	5	Q905Y6: anemonia su
33	180.5	14.9	233	5	Q905Y9: renilla ren
34	149.5	14.9	238	5	Q905Y0: ptilosarcus
35	187.5	14.7	227	5	Q905Y5: heteractis
36	183.5	14.4	231	5	Q905Y5: heteractis
37	179.5	14.1	227	5	Q905Y6: zoanthus sp
38	178	14.0	229	5	Q905Y8: condylactis
39	173.5	13.5	227	5	Q8T5E7: condylactis
40	172	13.3	231	5	Q905Y4: zoanthus sp
41	172	13.3	234	5	Q8T5F3: scyllamia cu
42	167.5	13.1	231	5	Q8T404: zoanthus sp
43	166.5	13.1	231	5	Q8T5E8: ricordea fl
44	156	12.2	231	5	Q8T5E8: discosoma s
45	149	11.7	231	5	Q8T5E9: ricordea fl

ALIGNMENTS

RESULT 1	Q27903	PRELIMINARY:	PRT:	238 AA
17	Q27903			
AC	Q27903			
BT	Q1-NOV-1996 (TRENBLREL_01, Created)			
DT	Q1-NOV-1996 (TRENBLREL_01, Last sequence update)			
DT	Q1-JUN-2002 (TRENBLREL_21, Last annotation update)			
DE	Green fluorescent protein:			
GN	GFP			
OS	unidentified			
SC	unclassified			
FX	NCBI:taxid-32644:			
AA	all			
BB	SEQUENCE FROM N.A.			
EX	MEDLINE:9729832: PubMed-9154931:			
KA	Kawakida G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.			
RT	Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage.			
KL	Plant Mol Biol 33:985-999(1997).			
EMBL	X96418: CAA65278.1			
EMBL	Q42122: CAA65278.1			
InterPro	IPR000786: Green_fl_protein			
PFAM	PF01353: GFP			
PRINTS	PRC1229: GFP			
Protein	Q905Y6: Green_fl_protein			
SEQUENCE	238 AA: CA932D47262AF2D3 CRC64:			

Query Match 99.74, Score 1271, DB 5, Length 238;
Res. Local Similarity 99.6%, Pred. No. 1.2e-97;
Matches 237: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MSKGEELFGVVPILVELDQVNGHKFSVSGEGGATYKULKFCITGK:FPWPPL 60	
DB	1	MKGEEELFGVVPILVELDQVNGHKFSVSGEGGATYKULKFCITGK:FPWPPL 60	
QY	61	VTTSYGVCFSRYPDHMKRRDFKFSAMPEGYVQERTIFFKDGNTKTRAEVKFGDILV 120	
DB	61	VTTSYGVCFSRYPDHMKRRDFKFSAMPEGYVQERTIFFKDGNTKTRAEVKFGDILV 120	


```

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240

RESULT 11
Q8WTC8 PRELIMINARY: PRT: 238 AA.
AC Q8WTC8:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
CS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=OPFXM1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435430; AAL33915;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27002 MW: 3558429822640018 CRC64;
Query Match 85.4%; Score 1089; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.5e-82;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240

RESULT 12
Q8WTC7 PRELIMINARY: PRT: 238 AA.
AC Q8WTC7:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Green fluorescent protein.

```

```

GN GFP.
CS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=OPFXM1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435430; AAL33915;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27002 MW: 3558429822640018 CRC64;
Query Match 85.4%; Score 1089; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 1.5e-82;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
DB 61 VTTESYVQVCFSPYVDHMKHGHGPKSAMPGGYVQERTIPFKDQSNYKTRAEVKEGDLV 120
QY 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
DB 121 NPTELKSGDKEDGNTLGHKLEYNFNSHNYIMADKQKNGIKVNFIRHNIEGSGVQLAD 180
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
QY 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240
DB 181 HYQNTTIGGGVLLPNDHNYLSQTSALSKDPNKRDMHVLLEFVTAAGTTHGMDELYK 240

RESULT 13
Q8WTC5 PRELIMINARY: PRT: 238 AA.
AC Q8WTC5:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
CS Aequorea macrodactyla.
CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
CC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=OPFXM1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen X.,
Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
macrodactyla."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435432; AAL33917;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR PRODOM: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA: 27018 MW: 75521EA5534E573A CRC64;
Query Match 85.3%; Score 1087; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 2.1e-82;
Matches 198; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKEFSVSGEGSDATYKLTKEICTTCKSLPVPWTL 60

```



```

SEQUENCE FROM N.A.
RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA Raikowski P., Gorbunov M., Kolber Z.;
RA *Green fluorescent proteins in Caribbean Scleractinian corals.*;
RA Submitted (Aug-2001) to the EMBL/GenBank/DDRC databases.
RA EMBL: AF384683; AAK62982.2;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; Grp; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE: 225 AA: 25847 MW: 77067DC616929AF CRC64:

Query Match          19.6%   Score 253.   DB 5: Length 225;
Seq Local Similarity 28.0%; Pred.No. 2.8e-13;
Matches    60; Conservative 42; Mismatches 84; Indels 25; Gaps -
QY      11 VVPILVLELDGWNCHKFSVSCEGGDATYCKLKIKELCTTG-KLPVKPKTLVTITSSVGVC 69
LB      7 IMEKLRKCGGVN:CHKKEVIKGEGKRFPGIQINLTVEGAPLFPAYDQLTSATGYGNKR 66
QY      70 QTSRYPHMKRKHDFKSAFNGEYVQCFTIFTKDGNNKTAEVKKFGEDFLVNRIKLKIGID 129
BB      67 VETKYDDIP-DYEKOTPEFGSWERIMAYEQSICIASTSIKKMGCDGFIEIQFHGVN 124
QY      130 FKESCNLGHK-LEYNVNSHNYIMADKQNKGIKNFKIKNIEDSGVGLA-----DHV 182
BB      125 FHPNGPVKKRLKWEFSTEKV-----VEDGVKGVNVNALLHGGRHY 169
QY      163 C---QNTPEGDSEVLIFDNHVLSTCSALSKDPNF 213
BB      170 RCHASTYKAKKAVQLDYHFVDHRRIILSHND 203

Search completed: July 15, 2003 14:13:24
Job time : 43.933 secs

```

Genome version 3.1.6
Copyright (c) 1995 - 2001, Genopharm Inc.

OR protein - protein search, using sw model

Run on: July 15, 2004, 16:11:38 : Search time in seconds
(without alignments)
574.63 Million cell updates/sec

Title: us-09-967-301-3
Patient score: 1226
Sequence: 1 MSKSEHFTGVPLVLELG.....VAGFVAGHTRZBHLK 234

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 6.5

Searched: 45189 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 45189

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA*

- 1: /cgn2.6/prodata/2/pubaa/US07_NEW_PUB.pep.*
- 2: /cgn2.6/prodata/2/pubaa/PC1_NEW_PUB.pep.*
- 3: /cgn2.6/prodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 5: /cgn2.6/prodata/2/pubaa/PC1_US_PUB.pep.*
- 6: /cgn2.6/prodata/2/pubaa/US07_PUB.pep.*
- 7: /cgn2.6/prodata/2/pubaa/US07_PUB.pep.*
- 8: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 9: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 10: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 11: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 12: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 13: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*
- 14: /cgn2.6/prodata/2/pubaa/US06_PUB.pep.*

Pred. No. is the number of residues predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	DB ID	Description
1	1258	98.6	238	9	US-09-967-301-3
2	1258	98.6	238	10	US-09-967-301-3
3	1258	98.6	238	12	US-10-024-684-2
4	1255	98.4	238	9	US-09-969-954B-2
5	1254	98.3	238	9	US-09-969-954B-2
6	1254	98.3	238	9	US-09-969-954B-2
7	1254	98.3	238	9	US-10-121-258-10
8	1254	98.3	238	9	US-10-121-258-10
9	1254	98.3	238	10	US-09-969-954B-2
10	1254	98.3	238	10	US-09-969-954B-2
11	1252	98.1	244	9	US-09-969-954B-2
12	1252	98.1	244	9	US-09-969-954B-2
13	1252	98.1	244	9	US-09-969-954B-2
14	1252	98.1	244	9	US-09-969-954B-2
15	1252	98.1	244	9	US-09-969-954B-2
16	1252	98.1	244	9	US-09-969-954B-2
17	1251	98.0	239	9	US-09-969-954B-2
18	1251	98.0	243	9	US-09-969-954B-2
19	1251	98.0	243	9	US-09-969-954B-2

1251	98.0	243	9	US-09-969-954B-2	Sequence 64, Appl
1251	98.0	243	9	US-09-969-954B-2	Sequence 66, Appl
1251	98.0	243	9	US-09-969-954B-2	Sequence 68, Appl
1251	98.0	243	9	US-09-969-954B-2	Sequence 70, Appl
1251	98.0	1070	9	US-10-001-486B-2	Sequence 2, Appl
1251	98.0	1099	9	US-10-259-864-4	Sequence 1, Appl
1251	98.0	1147	9	US-10-259-864-4	Sequence 2, Appl
1251	98.0	1452	12	US-10-050-673-2	Sequence 14, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 54, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 56, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 58, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 114, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 116, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 120, Appl
1250	98.0	243	9	US-09-969-954B-2	Sequence 122, Appl
1250	98.0	244	9	US-09-969-954B-2	Sequence 6, Appl
1250	98.0	244	9	US-09-969-954B-2	Sequence 20, Appl
1250	98.0	244	9	US-09-969-954B-2	Sequence 22, Appl
1249	97.9	243	9	US-09-969-954B-2	Sequence 4, Appl
1249	97.9	243	9	US-09-969-954B-2	Sequence 6, Appl
1249	97.9	243	9	US-09-969-954B-2	Sequence 8, Appl
1249	97.9	243	9	US-09-969-954B-2	Sequence 12, Appl
1249	97.9	243	9	US-09-969-954B-2	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-057-505-2
? Sequence 2, Application US/10057505
? Patent No. US2002015497A1
? GENERAL INFORMATION:
? APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
? APPLICANT: AURORA BIOSCIENCES CORPORATION
? APPLICANT: TSEH, Roger
? APPLICANT: HEM, Roger
? APPLICANT: CHITT, Andrew
? TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
? FILE REFERENCE: P0001260-3
? CURRENT APPLICATION NUMBER: US/10/057.505
? CURRENT FILING DATE: 2002-01-25
? PRIOR APPLICATION NUMBER: US 08/742,553
? PRIOR FILING DATE: 1997-01-21
? PRIOR APPLICATION NUMBER: US 09/396,003
? PRIOR FILING DATE: 1999-04-13
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: Patent 16 version 3.1
? SEQ ID NO 2
? LENGTH: 238
? TYPE: PPT
? ORGANISM: Aequorea victoria
US-10-057-505 2

Query Match	98.6%	Score 1258	DB 9	Length 238
Posit Local Similarity	98.3%	Pred. No. 60-115		
Matches 235	Conservative 0	Mismatches 3	Indels 1	Gaps 0
QY	1	MSKSEELFTGVPLVLELDGVNSHKPSVSGESDATYKLTLPICITCTSKLPVWPTL 66		
DB	1	MSKSEELFTGVPLVLELDGVNSHKPSVSGESDATYKLTLPICITCTSKLPVWPTL 60		
QY	61	VTLTSYVQGFSAVPMKPHDFKSAHPGYVCHWTIFKODGNYKTRAFKPEECLTV 120		
DB	61	VTFYSYVQGFSAVPMKPHDFKSAHPGYVCHWTIFKODGNYKTRAFKPEECLTV 120		
QY	121	NRIELKGLPFKESGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGVQLAD 180		
DB	121	NRIELKGLPFKESGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGVQLAD 180		

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238
 DB 161 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

RESULT 2

US-09-884-681-2

Sequence 2, Application US/09884681

Patent No. US20020061546A1

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

Cobalt, Andrex B.

TITLE OF INVENTION: Assays for Protein Kinases Using

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

STREET: Townsend and Townsend and Crew LLP

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-5834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/684,681

FILING DATE: 19-Jun-2001

CLASSIFICATION: Unknown

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/679,865

FILING DATE: Unknown

ATTORNEY/AGENT INFORMATION:

NAME: Storella, Joan S.

REFERENCE/DOCKET NUMBER: 023078-069000

TELEPHONE: (415) 576-6200

TELEFAX: (415) 576-6300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amine acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-884-681-2

Query Match 98.6%; Score 1258; DB 12; Length 238;

Best Local Similarity 98.7%; Pred. No. 60-115;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

DB 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

QY 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

DB 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

QY 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

DB 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

DB 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

RESULT 3

US-10-024-686-2

Sequence 2, Application US/098999548

Patent No. US20020157135A1

Query Match 98.6%; Score 1258; DB 12; Length 238;

Best Local Similarity 98.7%; Pred. No. 60-115;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

DB 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

QY 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

DB 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

QY 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

DB 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

DB 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

RESULT 4

US-09-899-954R-2

Sequence 2, Application US/098999548

Patent No. US20020157135A1

Query Match 98.6%; Score 1258; DB 12; Length 238;

Best Local Similarity 98.7%; Pred. No. 60-115;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

DB 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

QY 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

DB 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

QY 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

DB 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

DB 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

Sequence 2, Application US/10024686

Patent No. US2002012313A1

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

Heim, Roger

TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

STREET: Fife & Richardson P.C.

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/024,686

FILING DATE: 17-Dec-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/057,995

FILING DATE: Unknown

APPLICATION NUMBER: 04/727,452

FILING DATE: 10-Oct-1996

APPLICATION NUMBER: 0565/14692

FILING DATE: 13-Nov-1995

APPLICATION NUMBER: 05/337,915

FILING DATE: 10-Nov-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hailo, Lisa A.

REGISTRATION NUMBER: 33,347

REFERENCE/DOCKET NUMBER: 07267/0-2202

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/578-5070

TELEFAX: 619/674-5099

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amine acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-024-686-2

Query Match 98.6%; Score 1258; DB 12; Length 238;

Best Local Similarity 98.7%; Pred. No. 60-115;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

DB 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

QY 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

DB 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

QY 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

DB 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

DB 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

RESULT 4

US-09-899-954R-2

Sequence 2, Application US/098999548

Patent No. US20020157135A1

Query Match 98.6%; Score 1258; DB 12; Length 238;

Best Local Similarity 98.7%; Pred. No. 60-115;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

DB 1 MSKGEELFGVVPILVELGDVNGHKEFSVSGEGDATYKLTKEICTGKLPVWPPTL 60

QY 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

DB 61 VTLSVGVQCFSPYDPHMKHFFKSAHPGYVOERTIEFKDQGNKTKRAVKEEDTLV 120

QY 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

DB 121 NRTELKIDFKKGNILGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNIEDGVQLAD 180

QY 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

DB 181 HYQNTPIGPGVPLPNKHYSLSQALSQKDNKEXLHWLLGFFVTAAAGTTHGKDELYK 238

OTHER INFORMATION: Green fluorescent protein
US-09-852-036-1

Query Match 98.1%, Score 1254, DP 126, Length 244;
Best Local Similarity 98.1%, Pred. No. 2.4e-114;
Matches 234; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
QY 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
DB 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
QY 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
DB 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244

RESULT 11
US-09-899-954B-8
Sequence 9, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899-954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU P8077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Gen(CTA)6 GFP construct
US-09-899-954B-8

Query Match 98.1%, Score 1254, DP 126, Length 244;
Best Local Similarity 98.1%, Pred. No. 2.4e-114;
Matches 234; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
QY 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
DB 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
QY 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
DB 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244

RESULT 12
US-09-899-954B-10
Sequence 10, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899-954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU P8077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Gen(CTA)6 GFP construct
US-09-899-954B-10
Query Match 98.1%, Score 1254, DP 126, Length 244;
Best Local Similarity 97.9%, Pred. No. 2.4e-114;
Matches 233; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
DB 1 MSKGEELFTGVVPIVLELGGVNGHKFSVSGEGEDATYKLTAKFTCTGKLPVWPIL 60
QY 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
DB 61 VTLSYGVOCSRYPDHMKRKHDFKFSAMPEGVVQERTIFPKDDGNKPAFVKFEGLV 120
QY 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
DB 121 NRLEKGDGDFKEDGNILGHKLEYNVNSHNVYIMAKOKNGIKVNEKIRHNIEKGVGLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244
DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMYLLGVTAAGLTHGMDELYK 244

RESULT 13
US-09-899-954B-12
Sequence 12, Application US/09899954B
Patent No. US20020157135A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
FILE REFERENCE: 10338-60S
CURRENT APPLICATION NUMBER: US/09/899-954B
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU P8077
PRIOR FILING DATE: 1999-01-08
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 244
TYPE: PRT
ORGANISM: Artificial Sequence

```

FEATURE:
OTHER INFORMATION: Leu(CTT)6 GFP construct
US-09-899-954B-12

Query Match          98.1%  Score 1252:  DB 9:  Length 244:
Best Local Similarity 97.9%  Pred. No. 2.4e-114:
Matches 233:  Conservative 1:  Mismatches 4:  Indels 0:  Gaps 0:

CY 1 MSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
DB 7 LSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
CY 61 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
DB 67 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
CY 181 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 238
DB 187 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 244

RESULT 14
US-09-899-954B-14
Sequence 14: Application US/0989954B
Patent No. US20020157155A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
TITLE OF INVENTION: target cell or tissue of a plant
FILE REFERENCE: 10338-6J5
CURRENT APPLICATION NUMBER: US/09/899-954B
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: AU PP8C77
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PAT/AU00/0007
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 244
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Leu(CTT)6 GFP construct
US-09-899-954B-14

Query Match          98.1%  Score 1252:  DB 9:  Length 244:
Best Local Similarity 97.9%  Pred. No. 2.4e-114:
Matches 233:  Conservative 1:  Mismatches 4:  Indels 0:  Gaps 0:

CY 1 MSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
DB 7 LSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
CY 61 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
DB 67 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
CY 181 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 238
DB 187 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 244

```

```

RESULT 15
US-09-899-954B-16
Sequence 16: Application US/0989954B
Patent No. US20020157155A1
GENERAL INFORMATION:
APPLICANT: The University of Queensland (all designated States except US)
APPLICANT: Frazer, Ian Hector (US only)
APPLICANT: Zhou, Jian (US only)
APPLICANT: Botella, Jose (US only)
TITLE OF INVENTION: Polynucleotide and method for selectively expressing a protein
TITLE OF INVENTION: target cell or tissue of a plant
FILE REFERENCE: 10338-6J5
CURRENT APPLICATION NUMBER: US/09/899-954B
PRIOR FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: AU PP8C77
PRIOR FILING DATE: 1999-01-08
PRIOR APPLICATION NUMBER: PAT/AU00/0007
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 244
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Leu(CTT)6 GFP construct
US-09-899-954B-16

Query Match          98.1%  Score 1252:  DB 9:  Length 244:
Best Local Similarity 97.9%  Pred. No. 2.4e-114:
Matches 233:  Conservative 1:  Mismatches 4:  Indels 0:  Gaps 0:

CY 1 MSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
DB 7 LSKGEELFTGVVPIVLVDGVNGHKFSVSGEGSDATYGKLTKEFTCTGKLPVWPIL 60
CY 61 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
DB 67 VTTSYGVQCFSRYPDRMKRHDFEKSAMPEGVQVQERTIFFKDGNYKTRAEVKFEGDTLV 120
CY 121 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
DB 127 NRLELKGIDFKEDGNILGKLEYNNSHNYIMAKKNGIKVNFKIHNIHEDGVQGLAD 180
CY 181 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 238
DB 187 HYQONTPIGDGPVLDPNHYLSTQSALSKDPNEKRDHMLLEFVTAAGTTTHOMLELYK 244

```

Search completed: July 16, 2003, 16:30:10
 CPU time : 29 secs

GeneCore version 5.0.1.4
Copyright (c) 1994 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2003 15:17:12 : Search time at 0.67 seconds
(without alignment)
197.696 Million cell updates/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: 1 KSKKRLRTGVVIVLIVLGG.....VLIDFVAGDILHEDHMKK 249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9616422 residues

Total number of hits satisfying chosen parameters: 14224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match OK
Maximum Match 100%
Listing first 45 summaries

Database : PIR-73:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARY

Result	No.	Score	Match	Length	DB	ID	Description
1	1248	97.8	248	1	EC5514	1	green-fluorescent protein
2	107	8.4	785	1	H22428	2	hypothetical protein
3	99.5	7.8	287	2	B85592	2	hypothetical protein
4	92.5	7.2	534	1	EC1014	1	nitrogenase (EC 1.1.1.1)
5	90	7.1	1224	1	FRH548	1	nitrogenase (EC 1.1.1.1)
6	89.5	6.1	861	2	H64122	2	nitrogenase (EC 1.1.1.1)
7	88.5	6.9	370	2	E70390	2	nitrogenase (EC 1.1.1.1)
8	86	6.9	353	2	B84341	2	nitrogenase (EC 1.1.1.1)
9	87.5	6.9	792	2	G24278	2	nitrogenase (EC 1.1.1.1)
10	87.5	6.9	804	2	E64102	2	nitrogenase (EC 1.1.1.1)
11	87.5	6.9	222	1	A56228	1	DNA-directed RNA polymerase
12	87	6.8	357	2	G61355	2	DNA-directed RNA polymerase
13	87	6.8	432	2	GH414	2	synthetase (EC 1.1.1.1)
14	86.5	6.8	263	2	S53488	2	water stress inducible
15	86.5	6.8	860	2	AC5582	2	leucyl tRNA synthetase
16	86.5	6.8	257	2	E74514	2	hypothetical protein
17	85.5	6.7	655	2	D83017	2	DNA topoisomerase
18	85	6.7	612	2	T06596	2	DNA binding protein
19	84.5	6.4	730	1	C14092	1	calpain (EC 3.4.22.1)
20	84.5	6.6	889	2	JC5576	2	inter-alpha-trypsin inhibitor
21	84	6.6	281	2	AC2052	2	hypothetical protein
22	84	6.6	51	2	C95338	2	hypothetical protein
23	84	6.6	874	2	JC4336	2	serine protease
24	83.5	6.5	613	2	A95552	2	oligodeoxyribonuclease
25	83	6.5	578	1	I40734	1	thiolipase
26	82.5	6.5	312	2	C61710	2	thiolipase
27	82.5	6.5	422	2	B24515	2	calpain (EC 3.4.22.1)
28	82.5	6.5	502	2	B90103	2	thiolipase
29	82.5	6.5	941	2	S29443	2	calpain (EC 3.4.22.1)

30	82.5	6.5	1134	2	A60234	2	IgA E0 receptor pr
31	82.5	6.5	1154	1	F05049	1	IgA E0 receptor pr
32	82	6.4	470	2	T33049	2	hypothetical prote
33	82	6.4	774	2	T33049	2	alpha-amylase homo
34	82	6.4	855	2	T11852	2	lipoxigenase (EC 1
35	81.5	6.4	310	2	S68225	2	synthetase (EC 1.1
36	81.5	6.4	425	2	C97354	2	hypothetical prote
37	81.5	6.4	564	2	E81317	2	ABC type transport
38	81.5	6.4	788	1	J0V1H8	1	DNA-directed DNA p
39	81	6.3	336	2	G64459	2	hypothetical prote
40	81	6.3	351	2	A88403	2	protein W3011.7.1
41	81	6.3	451	2	T06536	2	photosystem II chl
42	81	6.3	1516	2	T41235	2	proteobacteriophage
43	80.5	6.3	989	2	S84245	2	NADH dehydrogenase
44	80.5	6.3	893	2	G64613	2	conserved tyrosine
45	80.5	6.3	553	1	A11535	1	trypsin-like phosph

ALIGNMENTS

RESULT 1

JO1514
green-fluorescent protein [validated] - Hydromedusa (Aequorea victoria)
C:Species: Aequorea victoria
C:Date: 03-Dec-1999 #sequence, revision: 03-Dec-1999 #text change 23-Mar-2001
C:Accession: J01514; P03335; S48693; S51330; S51331
R:Fraser, D.C.; Eckhardt, V.K.; Ward, W.W.; Prendergast, F.G.; Gorman, M.J.
Gene 111, 229-233, 1992
A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A:Reference number: JO1514; MUID:92175527; PMID:1347477
A:Accession: J01514
A:Molecule type: DNA
A:Residues: 1-107, 'S', 109-232 <PRAL>
A:Cross-references: GR:M62654; NID:q155662; PDB:AAA27722.1; PDB:q155663
A:Accession: JC5114
A:Molecule type: mRNA
A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>
A:Cross-references: GR:M62653; NID:q155660; PDB:AAA27721.1; PDB:q155661
A:Accession: PQ0335
A:Molecule type: protein
A:Residues: 46-64, 77-122, 132-151, 154 183-185-200 <PRA3>
PDB:1A70, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, S36, S37, S38, S39, S40, S41, S42, S43, S44, S45, S46, S47, S48, S49, S50, S51, S52, S53, S54, S55, S56, S57, S58, S59, S60, S61, S62, S63, S64, S65, S66, S67, S68, S69, S70, S71, S72, S73, S74, S75, S76, S77, S78, S79, S80, S81, S82, S83, S84, S85, S86, S87, S88, S89, S90, S91, S92, S93, S94, S95, S96, S97, S98, S99, S100, S101, S102, S103, S104, S105, S106, S107, S108, S109, S110, S111, S112, S113, S114, S115, S116, S117, S118, S119, S120, S121, S122, S123, S124, S125, S126, S127, S128, S129, S130, S131, S132, S133, S134, S135, S136, S137, S138, S139, S140, S141, S142, S143, S144, S145, S146, S147, S148, S149, S150, S151, S152, S153, S154, S155, S156, S157, S158, S159, S160, S161, S162, S163, S164, S165, S166, S167, S168, S169, S170, S171, S172, S173, S174, S175, S176, S177, S178, S179, S180, S181, S182, S183, S184, S185, S186, S187, S188, S189, S190, S191, S192, S193, S194, S195, S196, S197, S198, S199, S200, S201, S202, S203, S204, S205, S206, S207, S208, S209, S210, S211, S212, S213, S214, S215, S216, S217, S218, S219, S220, S221, S222, S223, S224, S225, S226, S227, S228, S229, S230, S231, S232, S233, S234, S235, S236, S237, S238, S239, S240, S241, S242, S243, S244, S245, S246, S247, S248, S249, S250, S251, S252, S253, S254, S255, S256, S257, S258, S259, S260, S261, S262, S263, S264, S265, S266, S267, S268, S269, S270, S271, S272, S273, S274, S275, S276, S277, S278, S279, S280, S281, S282, S283, S284, S285, S286, S287, S288, S289, S290, S291, S292, S293, S294, S295, S296, S297, S298, S299, S300, S301, S302, S303, S304, S305, S306, S307, S308, S309, S310, S311, S312, S313, S314, S315, S316, S317, S318, S319, S320, S321, S322, S323, S324, S325, S326, S327, S328, S329, S330, S331, S332, S333, S334, S335, S336, S337, S338, S339, S340, S341, S342, S343, S344, S345, S346, S347, S348, S349, S350, S351, S352, S353, S354, S355, S356, S357, S358, S359, S360, S361, S362, S363, S364, S365, S366, S367, S368, S369, S370, S371, S372, S373, S374, S375, S376, S377, S378, S379, S380, S381, S382, S383, S384, S385, S386, S387, S388, S389, S390, S391, S392, S393, S394, S395, S396, S397, S398, S399, S400, S401, S402, S403, S404, S405, S406, S407, S408, S409, S410, S411, S412, S413, S414, S415, S416, S417, S418, S419, S420, S421, S422, S423, S424, S425, S426, S427, S428, S429, S430, S431, S432, S433, S434, S435, S436, S437, S438, S439, S440, S441, S442, S443, S444, S445, S446, S447, S448, S449, S450, S451, S452, S453, S454, S455, S456, S457, S458, S459, S460, S461, S462, S463, S464, S465, S466, S467, S468, S469, S470, S471, S472, S473, S474, S475, S476, S477, S478, S479, S480, S481, S482, S483, S484, S485, S486, S487, S488, S489, S490, S491, S492, S493, S494, S495, S496, S497, S498, S499, S500, S501, S502, S503, S504, S505, S506, S507, S508, S509, S510, S511, S512, S513, S514, S515, S516, S517, S518, S519, S520, S521, S522, S523, S524, S525, S526, S527, S528, S529, S530, S531, S532, S533, S534, S535, S536, S537, S538, S539, S540, S541, S542, S543, S544, S545, S546, S547, S548, S549, S550, S551, S552, S553, S554, S555, S556, S557, S558, S559, S560, S561, S562, S563, S564, S565, S566, S567, S568, S569, S570, S571, S572, S573, S574, S575, S576, S577, S578, S579, S580, S581, S582, S583, S584, S585, S586, S587, S588, S589, S590, S591, S592, S593, S594, S595, S596, S597, S598, S599, S600, S601, S602, S603, S604, S605, S606, S607, S608, S609, S610, S611, S612, S613, S614, S615, S616, S617, S618, S619, S620, S621, S622, S623, S624, S625, S626, S627, S628, S629, S630, S631, S632, S633, S634, S635, S636, S637, S638, S639, S640, S641, S642, S643, S644, S645, S646, S647, S648, S649, S650, S651, S652, S653, S654, S655, S656, S657, S658, S659, S660, S661, S662, S663, S664, S665, S666, S667, S668, S669, S670, S671, S672, S673, S674, S675, S676, S677, S678, S679, S680, S681, S682, S683, S684, S685, S686, S687, S688, S689, S690, S691, S692, S693, S694, S695, S696, S697, S698, S699, S700, S701, S702, S703, S704, S705, S706, S707, S708, S709, S710, S711, S712, S713, S714, S715, S716, S717, S718, S719, S720, S721, S722, S723, S724, S725, S726, S727, S728, S729, S730, S731, S732, S733, S734, S735, S736, S737, S738, S739, S740, S741, S742, S743, S744, S745, S746, S747, S748, S749, S750, S751, S752, S753, S754, S755, S756, S757, S758, S759, S760, S761, S762, S763, S764, S765, S766, S767, S768, S769, S770, S771, S772, S773, S774, S775, S776, S777, S778, S779, S780, S781, S782, S783, S784, S785, S786, S787, S788, S789, S790, S791, S792, S793, S794, S795, S796, S797, S798, S799, S800, S801, S802, S803, S804, S805, S806, S807, S808, S809, S810, S811, S812, S813, S814, S815, S816, S817, S818, S819, S820, S821, S822, S823, S824, S825, S826, S827, S828, S829, S830, S831, S832, S833, S834, S835, S836, S837, S838, S839, S840, S841, S842, S843, S844, S845, S846, S847, S848, S849, S850, S851, S852, S853, S854, S855, S856, S857, S858, S859, S860, S861, S862, S863, S864, S865, S866, S867, S868, S869, S870, S871, S872, S873, S874, S875, S876, S877, S878, S879, S880, S881, S882, S883, S884, S885, S886, S887, S888, S889, S890, S891, S892, S893, S894, S895, S896, S897, S898, S899, S900, S901, S902, S903, S904, S905, S906, S907, S908, S909, S910, S911, S912, S913, S914, S915, S916, S917, S918, S919, S920, S921, S922, S923, S924, S925, S926, S927, S928, S929, S930, S931, S932, S933, S934, S935, S936, S937, S938, S939, S940, S941, S942, S943, S944, S945, S946, S947, S948, S949, S950, S951, S952, S953, S954, S955, S956, S957, S958, S959, S960, S961, S962, S963, S964, S965, S966, S967, S968, S969, S970, S971, S972, S973, S974, S975, S976, S977, S978, S979, S980, S981, S982, S983, S984, S985, S986, S987, S988, S989, S990, S991, S992, S993, S994, S995, S996, S997, S998, S999, S1000, S1001, S1002, S1003, S1004, S1005, S1006, S1007, S1008, S1009, S1010, S1011, S1012, S1013, S1014, S1015, S1016, S1017, S1018, S1019, S1020, S1021, S1022, S1023, S1024, S1025, S1026, S1027, S1028, S1029, S1030, S1031, S1032, S1033, S1034, S1035, S1036, S1037, S1038, S1039, S1040, S1041, S1042, S1043, S1044, S1045, S1046, S1047, S1048, S1049, S1050, S1051, S1052, S1053, S1054, S1055, S1056, S1057, S1058, S1059, S1060, S1061, S1062, S1063, S1064, S1065, S1066, S1067, S1068, S1069, S1070, S1071, S1072, S1073, S1074, S1075, S1076, S1077, S1078, S1079, S1080, S1081, S1082, S1083, S1084, S1085, S1086, S1087, S1088, S1089, S1090, S1091, S1092, S1093, S1094, S1095, S1096, S1097, S1098, S1099, S1100, S1101, S1102, S1103, S1104, S1105, S1106, S1107, S1108, S1109, S1110, S1111, S1112, S1113, S1114, S1115, S1116, S1117, S1118, S1119, S1120, S1121, S1122, S1123, S1124, S1125, S1126, S1127, S1128, S1129, S1130, S1131, S1132, S1133, S1134, S1135, S1136, S1137, S1138, S1139, S1140, S1141, S1142, S1143, S1144, S1145, S1146, S1147, S1148, S1149, S1150, S1151, S1152, S1153, S1154, S1155, S1156, S1157, S1158, S1159, S1160, S1161, S1162, S1163, S1164, S1165, S1166, S1167, S1168, S1169, S1170, S1171, S1172, S1173, S1174, S1175, S1176, S1177, S1178, S1179, S1180, S1181, S1182, S1183, S1184, S1185, S1186, S1187, S1188, S1189, S1190, S1191, S1192, S1193, S1194, S1195, S1196, S1197, S1198, S1199, S1200, S1201, S1202, S1203, S1204, S1205, S1206, S1207, S1208, S1209, S1210, S1211, S1212, S1213, S1214, S1215, S1216, S1217, S1218, S1219, S1220, S1221, S1222, S1223, S1224, S1225, S1226, S1227, S1228, S1229, S1230, S1231, S1232, S1233, S1234, S1235, S1236, S1237, S1238, S1239, S1240, S1241, S1242, S1243, S1244, S1245, S1246, S1247, S1248, S1249, S1250, S1251, S1252, S1253, S1254, S1255, S1256, S1257, S1258, S1259, S1260, S1261, S1262, S1263, S1264, S1265, S1266, S1267, S1268, S1269, S1270, S1271, S1272, S1273, S1274, S1275, S1276, S1277, S1278, S1279, S1280, S1281, S1282, S1283, S1284, S1285, S1286, S1287, S1288, S1289, S1290, S1291, S1292, S1293, S1294, S1295, S1296, S1297, S1298, S1299, S1300, S1301, S1302, S1303, S1304, S1305, S1306, S1307, S1308, S1309, S1310, S1311, S1312, S1313, S1314, S1315, S1316, S1317, S1318, S1319, S1320, S1321, S1322, S1323, S1324, S1325, S1326, S1327, S1328, S1329, S1330, S1331, S1332, S1333, S1334, S1335, S1336, S1337, S1338, S1339, S1340, S1341, S1342, S1343, S1344, S1345, S1346, S1347, S1348, S1349, S1350, S1351, S1352, S1353, S1354, S1355, S1356, S1357, S1358, S1359, S1360, S1361, S1362, S1363, S1364, S1365, S1366, S1367, S1368, S1369, S1370, S1371, S1372, S1373, S1374, S1375, S1376, S1377, S1378, S1379, S1380, S1381, S1382, S1383, S1384, S1385, S1386, S1387, S1388, S1389, S1390, S1391, S1392, S1393, S1394, S1395, S1396, S1397, S1398, S1399, S1400, S1401, S1402, S1403, S1404, S1405, S1406, S1407, S1408, S1409, S1410, S1411, S1412, S1413, S1414, S1415, S1416, S1417, S1418, S1419, S1420, S1421, S1422, S1423, S1424, S1425, S1426, S1427, S1428, S1429, S1430, S1431, S1432, S1433, S1434, S1435, S1436, S1437, S1438, S1439, S1440, S1441, S1442, S1443, S1444, S1445, S1446, S1447, S1448, S1449, S1450, S1451, S1452, S1453, S1454, S1455, S1456, S1457, S1458, S1459, S1460, S1461, S1462, S1463, S1464, S1465, S1466, S

A>Title: HEP-COP, a novel human gene whose product is highly homologous to the alpha
A:Reference number: JC468; MID:96194806; PID:B64751
A:Accession: JC468
A:Molecule type: mRNA
A:Residues: 1-1224 <CHO>
A:Cross-references: GB-U24105; NID:q1636674; PION:AAB70379.1; PID:g162366
A:Experimental source: HepB hepatocellular carcinoma cell
A:Feuille, G.E.; Hanscher, G.; Kuslik, R.; Meyer, H.E.; Mettger, J.W.
J. Biol. Chem. 267, 22305-22309, 1992
A>Title: Identification of xenin, a xenopsin-related peptide, in the human gastric mu
A:Reference number: A44317; MID:94254515; PID:1429581
A:Accession: A44317
A:Molecule type: protein
A:Residues: 1-25 <FEQ>
A:Experimental source: gastric mucosa
A>Note: sequence extracted from NCBI backbone (NCBP:117913)
A>Note: plasma levels of xenin 2%, as determined by immunoassay, rise after meals and
C:Genetics:
A:Gene: GJB-GCPA; HEP-COP
A:Cross-references: GD6:4642787; OMIM:601924
A:Map position: 1q23-q25
C:Superfamily: yeast coatomer complex alpha chain; WD repeat homology
C:Keywords: duplication; hormone; plasma; stomach
F:1-25/Product: xenin; 25 *status experimental <XNP>
F:5-38/Domain: WD repeat homology <WD1>
F:17-25/Product: xenopsin-related peptide *status predicted <XNP>
F:17-80/Domain: WD repeat homology <WD2>
F:89-122/Domain: WD repeat homology <WD3>
F:131-164/Domain: WD repeat homology <WD4>
F:201-234/Domain: WD repeat homology <WD5>
F:245-278/Domain: WD repeat homology <WD6>

Query Match 7.1% Score 90 DB 1: Length 1224:
Best Local Similarity 25.9%; Pred. No. 23:
Matches 41; Conservative 26; Mismatches 59; Indexes 32; Gaps 7:

QY 90 EGYVQRITFFKDDSNKYRAEVKPESDTIVRIE-LKGIDFKEDGNILCHKLEYWNSH 148
 ||| | ||| ||||| :| :| :| :| :| :| :| :|
Db 612 KGVFEVALHFVDKE--KTRSLALECSNIEIALEAKALDKNCKWEKLGEVALGSGNHQ 668

QY 149 NVYMADKRGNGIKNVF-----KIRHNEDEGVO--LAHYQNTPTENGSPVLPPD 197
 ||| | ||| ||||| :| :| :| :| :| :| :| :|
Db 689 IVEMCYQR-KNFDKVSFLYLSTONLKKRKMKIATIRKMDSGHYGNALYSR----- 742

QY 198 NHYLSLTQSALSNDENKEDIHVLLGFVTAAAGIHGMCE 235
 ||| | ||| ||||| :| :| :| :| :| :| :| :|
Db 742 ---VSESVILKWGGKS-----LVCTAA---THGLDE 769

RESULT #
#64102
leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Ed KW20)
N:Alternate names: leucyl-tRNA synthetase
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 *sequence_revision 16-Aug-1995 *ext_change 03-Jun-2002
C:Accession: H64102
P:Fritschmann, R.D.; Adams, M.D.; White, C.; Clayton, R.A.; Kirkness, E.F.; Kerlavac
S.; Gouyave, J.B.; Scott, J.L.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Waldman
P.; F.M.; Brandon, R.C.; Finn, T.D.; Fitzchum, J.L.; Fuhrmann, J.L.; Geophagen, N.S.H.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, I.A.; Small, K.V.; Fraser, C.M.; Smith, H.G.; Venter
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64009; MID:95359636; PMID:7542860
A:Accession: H64102
A>Status: nucleic acid
A:Molecule type: DNA
A:Residues: 1-36<*TIGR>
A:Cross-references: GB-U32774; GH:L42223; NID:q1573942; PID:AA0225el.1; PID:q1573943
C:Genetics:
A:Gene: leus
C:Superfamily: leucine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 7.08; Score 89.5; DB 2; Length 561;
 Best Local Similarity 24.3%; Pred. No. 16;
 Matches 45; Conservative 26; Mismatches 67; Indels 47; Gaps 3;
 QY 49 TTGKLPVWPTLVITLSGVGQCFSSYPDHRKRDFFKSNMPGEGYVQERTFFKDDGNKY; 108
 DB 374 TCDKLPV-KVANFVLMHYGAGVAVPAHQDQ--FFAGKYSLFKQV-APLADLEIDL 370
 QY 109 KAEVKEGDTLVNFKELGIDFKEDGNILGHKLEFYNNYNSRVYLNADGKN 159
 DB 371 TKQAFVERGKLVNDEFGKNE--DGAFNG-----TADKLEKLVGSKHQVN 414
 QY 165 FKIRH-----NIEDGGVGLADHYQONFEGGQGVLLPDNHYL-STQSA; 207
 DB 415 YRLDQGVSRORYGAPIMCLLENGDVPA-----FMEDLPILPDVYVMDVNSPI 467
 QY 208 SKDPN 212
 DB 466 NADPN 472
 RESULT 7
 E70360
 iron-sulfur cofactor synthesis protein nifs; Aquifex aeolicus
 N:Contains: L-cysteine sulfoxidtransferase (EC 2.8.1.1);
 C:Species: Aquifex aeolicus
 C>Date: 08-May-1998 #sequence_revision 08-May-1998 #tax_change 07-Dec-1998
 C:Accession: E70360
 R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Jones, A.L.; Graham, D.E.; V.
 Nature 392, 354-359, 1998
 A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70360; MIM:36196566; PMID:9543620
 A:Accession: E70360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-370 <SIC>
 A:Cross-references: GB:A2000720; NID:42983529; PID:AA007111; PID:2848546; GB:A390045
 A:Experimental source: strain V55
 C:Genetics:
 C:Superfamily: nitrogen fixation protein nifs
 C:Keywords: phosphoprotein; pyridoxal phosphate; sulfurtransferase
 F195/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted
 F318/Active site: Cys (cysteine persulfide intermediate) #status predicted
 Query Match 6.9%; Score 84.5; DB 2; Length 370;
 Best Local Similarity 25.4%; Pred. No. 6.4;
 Matches 49; Conservative 30; Mismatches 91; Indels 26; Gaps 8;
 QY 3 KGEELFTGW----PILVELD---QGVNHRK-SVSGES--DEKATKCKLTAEFCT 49
 DB 104 KQVLLDVAVALGKPIEKANIVATSGKPKPAKGSFLYISGANVETPLIVSGQE 223
 QY 50 TKIP-----VWMTLVITLSGVGQCFSSYPDHRKRDFFKSNMPGEGYVQERTFFK 109
 DB 224 NGKSGTENWVGLLSAKALETIVSNESKYQEQKLRDGFENLELEA-LFEASIVGKDA 262
 QY 104 GNYKTRAEV---KREGDTLVNFKELGIDFKEDGNILGHKLEFYNNYNSRVYLNADGKN 159
 DB 283 ERSPISSIVMPKFGATVANKLSEKGYCSTGSACTSGEYEPKHKMLKMGFSGEKALAM 342
 QY 161 IKVNEKIRHNTFD 173
 DB 343 VRFSGILNKEEF 355
 RESULT 8
 E84941
 imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) [reported]; B.clinera sp. (strain
 E84941)
 C:Species: Buchnera sp.
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #tax_change 23-Mar-2001
 C:Accession: E84941

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 61-66, 2000
 A>Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
 A:Reference number: A84940; MIM:20445173; PMID:10993077
 A:Accession: E84941
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-353 <SIC>
 A:Cross-references: GB:AF000399; GSPDB:GN02144
 A:Experimental source: strain APS
 C:Genetics:
 A:Gene: hlsB; hsl02
 C:Superfamily: hlsB bifunctional enzyme; imidazoleglycerol-phosphate dehydratase homo
 C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 6.9%; Score 88; DB 2; Length 353;

Best Local Similarity 24.3%; Pred. No. 6.6;
 Matches 49; Conservative 33; Mismatches 79; Indels 42; Gaps 10;

QY 63 TLSYGVGQCFSSYPDHRKRDFFKSNMPGEGYVQERTIF---FKDDGNKYKTRAEVKE----- 114
 DB 62 TESEPLQDPS--TAEFLMLSVFS---EGVIFDQCLGPHFLDDUCVGRKPKIKRTEPWL 115
 QY 115 -----EKDQ-----LVNFKELGIDFKEDGNILGHKLEFYNNYNSRVYLNADGKN 159
 DB 116 DKLILKSSVIGDRTDQGLSNLAKIKCKYKEDCNMLHITKY-LKHNAYAILIETK 174
 QY 160 GYKVFKEHNIE-----DGGVGLADHYQONTPIGSGVLLPDNHYLSTQSAISKDFNEK 214
 DB 175 ETKAKIKVLMLEETSKIDGVKFFDMLHQLSIVSSGICM-----NISQGLDLDIDUHT 229
 QY 215 -RHRMVLGQ--FVTAAGITGCM 233
 DB 230 IETGIVLGEALQALGKNGK 251

RESULT 9

JC4078
 protective surface antigen D-15 precursor - Haemophilus influenzae (type D)
 C:Species: Haemophilus influenzae
 A:Variety: type D
 C>Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #tax_change 29-Sep-1999
 C:Accession: JC4078
 R:Black, F.S.; Loomis, S.; Chong, F.; Thomas, W.K.
 Gene 156, 97-99, 1995
 A>Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus I
 A:Reference number: JC4078; MIM:35255876; PMID:777523
 A:Accession: JC4078
 A:Molecule type: DNA
 A:Residues: 1-797 <SIC>
 A:Cross-references: GB:U13561; NID:3537447; PID:AAA85045.1; PID:353744b
 A:Experimental source: type b
 C:Superfamily: protective surface antigen D-15
 C:Keywords: surface antigen
 F119/Domains: signal sequence #status predicted <SIC>
 F120-797/Product: protective surface antigen D-15 #status predicted <MAI>

Query Match 6.9%; Score 67.5; DB 2; Length 797;
 Best Local Similarity 22.4%; Pred. No. 21;

Matches 49; Conservative 34; Mismatches 73; Indels 54; Gaps 12;
 QY 64 TSVGVGQCFSSYPDHRKRDFF-----FKSAMIEGYVQF-----PTI 98
 DB 427 TGVGTESGISYGSASVKQCNFLDTGAASVAGTCKNDYGTGVNLGYTETPYFKGVSLGGNV 486
 QY 39 FFKDDGNKYKTRAEVKEGDTLVNFKELGIDFKEDGNILGHKLEFYNNYNSRVYLNADGKN 147
 DB 487 FFEVDSKSDTSSTSNKYKRTIYGSNTILCFPVNENNSYVGLGTYNKLISNPALEYN--- 542
 QY 148 HNVYIMADKQK-NGKVN-----FKIRH-----IEDGGVGLADHYQONTPIGSGVLL 195
 DB 543 RNLYTOSMKFGKNGIKTINDFDFSGWNNYNSLNRGYEPTKGVKAS-----LG-GRVTI 593

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 16:01:23 : Search time 15.6672 seconds
(without alignments)
325.431 Million byte updates/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: MSKGFELTRVVPFLVLELDG.....VLNAVYVAVNTVWFLVWLVKLVK

Scoring table: BLASTNM2

Gapop 10.0 : Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum fb seq length: 3

Maximum DB seq length: 250000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Clustering first 45 summaries

Database : SwissProt_40:*

Freq. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1254	98.3	238	1	GFP_ARQV1	P12212	Aequorea vi
2	99.5	7.6	879	1	SYL_XITFA	Q9P6D6	Xyrella fas
3	92.5	7.2	533	1	NIPD_GLOPA	P04167	Caenorhabditis
4	90	7.1	1224	1	COPA_HUMAN	P53621	Human suprad
5	69.5	7.0	851	1	SYL_HAEIN	P33827	Caenorhabditis
6	84	6.9	353	1	R157_HYCAL	P87299	Buchnera ap
7	87.5	6.8	755	1	D152_HAEIN	P44935	Caenorhabditis
8	87.5	6.3	797	1	D152_HAEIN	P44934	Caenorhabditis
9	87.5	6.9	1224	1	COPA_HUMAN	P53621	Human suprad
10	87.5	6.9	2222	1	DEFA_YEAST	P24951	Saccharomy
11	87	6.8	357	1	TRNA_CAMEF	Q66982	Campylobact
12	87	6.8	439	1	SYN2_HIS3X	P24566	Discotheca
13	85.5	6.8	793	1	D152_HAEIN	P44935	Caenorhabditis
14	84.5	6.6	700	1	CAN2_HUMAN	P17655	Human sapro
15	84.5	6.6	896	1	TH13_HESAD	P97280	Desulfotol
16	84	6.6	874	1	SCAP_BACUL	P44952	Caenorhabditis
17	83.5	6.5	813	1	PRPF_XPCPU	Q98690	Myoglyasma
18	82.5	6.5	812	1	TRXB_CHEMU	P29817	Chlamydia
19	82.5	6.5	824	1	CAN2_PIG	P43187	Sus scrofa
20	82.5	6.5	822	1	CAN2_HA517	P55814	Gracilicoccus
21	82.5	6.5	659	1	ADAN_HUMAN	Q00116	Human sapro
22	82.5	6.5	941	1	GDA_HESAD	P24424	Caenorhabditis
23	82.5	6.5	1344	1	RAD1_STRAI	P27351	Streptococ
24	82	6.4	774	1	AMP2_SCHRC	P42419	Schistosom
25	81.5	6.4	533	1	CP51_CANSA	P55869	Canis lupus
26	81.5	6.4	538	1	GRDF_RAT	P84900	Rattus norv
27	81.5	6.4	798	1	YD4L_HRPHE	P13845	Herpes betat
28	81	6.3	336	1	YD4L_HRPHE	Q98743	Methanococ
29	81	6.4	441	1	PSB1_CVATA	P45104	Cyanobactera
30	80.5	6.3	333	1	MT04_HCFPV	P25443	Helicobacte
31	80.5	6.3	700	1	CAN2_MOUSE	P53929	Mus muscul
32	80.5	6.3	752	1	NKCL_RAT	P28840	Rattus norv
33	80	6.3	953	1	YNM7_YEAST	P53917	Saccharomy

RESULT 1
GFP_ARQV1
ID: GFP_ARQV1 STANDARD: PRI: 238 AA.

AC: P42212: 017104:
DE: 01-NOV-1995 (Rel: 32, Created)

DI: 01-NOV-1995 (Rel: 32, Last sequence update)

DI: 16-OCT-2001 (Rel: 40, Last annotation update)

DE: Green fluorescent protein.

GN: GFP.

OS: Aequorea victoria (Jellyfish).

OC: Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroida; Leptomedusae;

OC: Aequoreidae; Aequorea.

OX: NCBI_TaxID:5100;

RN: 1;

RP: SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX: MEDLINE:92175527; PubMed:1347277;

RA: Prasher B.G., Eckenrode V.K., Ward W.W., Prendergast F.G.,

RA: Cormier M.J.;

RT: "Primary structure of the Aequorea victoria green-fluorescent

RT: protein.";

RT: Gene 111:229-233(1992).

RN: 1;

RP: SEQUENCE FROM N.A.

RX: MEDLINE:94185810; PubMed:6147953;

RA: Iizuye S., Tsuji F.;

RT: "Aequorea green fluorescent protein. Expression of the gene and

RT: fluorescence characteristics of the recombinant protein.";

RL: FEBS Lett. 341:277-280(1994).

RP: GENE:PROTE.

RX: MEDLINE:93192221; PubMed:8448132;

RA: Cooch G.W., Prasher B.G., Westler W.M., Prendergast F.G., Ward W.W.;

RT: "Chemical structure of the hexapeptide chromophore of the Aequorea

RT: green-fluorescent protein.";

RL: Biochemistry 32:1212-1226(1993).

RP: X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX: MEDLINE:96355665; PubMed:8703025;

RA: Grime M., Cobitt A.B., Kallio K., Gross L.A., Tsien R.Y.,

RA: Reibman S.J.;

RT: "Crystal structure of the Aequorea victoria green fluorescent

RT: protein.";

RL: Science 273:1392-1395(1996).

RP: X-RAY CRYSTALLOGRAPHY (1.0 ANGSTROMS).

RX: MEDLINE:98294541; PubMed:9631087;

RA: Yang F., Moss L.G., Phillips G.N. Jr.;

RT: "The molecular structure of green fluorescent protein.";

RL: Nat. Biotechnol. 14:1246-1251(1996).

RP: X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.

RX: MEDLINE:98455509; PubMed:9782051;

RA: Wacker R.M., Eisliger M.A., Kallio K., Hanson G.T., Remington S.J.;

RT: "Structural basis of spectral shifts in the yellow-emission variants

RT: of green fluorescent protein.";

RL: Structure 6:1267-1277(1998).

ALIGNMENTS

CC AND FETAL TISSUES, XENIN IS FOUND IN GASTRIC, ENDOCRINE AND CENTRAL
CC MUCOSA. CIRCULATES IN THE BLOOD. SEEMS TO BE CONFINED TO SECRETORY
CC ENDOCRINE CELLS.
CC -1- DEVELOPMENTAL STAGE: XENIN IS RELEASED INTO THE CIRCULATION AFTER
CC A MEAL.
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (REF:ASP 1994S).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL institution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch)
CC
CC EMBL: U24105; AAB70879.1; 1;
CC Genew: HGNC:2230; CABA.
CC XIM: 601924; 1;
CC InterPro: IPR001490; WD40.
CC Pfam: PF00409; WD40; 7.
CC PRINTS: PR00130; GPROTEINRBP1.
CC ProDom: PD000018; WD40; 6.
CC SMART: SM00320; WD40; 6.
CC PROSITE: PS00676; WD REPEATS_1; 1.
CC PROSITE: PS00082; WD REPEATS_2; 6.
CC PROSITE: PS00294; WD REPEATS_REGION; 1.
CC Transport: Protein transport: Golgi stack; Endoplasmic reticulum;
KW Membrane: Repeat: WD repeat; Hormone.
FT PEPTIDE 1 25 XENIN.
FT PEPTIDE 1 35 PRORENIN.
FT REPEAT 7 47 WD 1.
FT REPEAT 49 79 WD 2.
FT REPEAT 93 121 WD 3.
FT REPEAT 133 163 WD 4.
FT REPEAT 203 233 WD 5.
FT REPEAT 247 277 WD 6.
SQ SEQUENCE 1224 AA: 138331 MW: 5A3B08R4D5B8F15D CRC64.

Query Match 7.1%; Score 90; DB 1; Length 1224;
Best Local Similarity 25.9%; Pred. No. 7.3;
Matches 41; Conservative 26; Mismatches 59; Indels 32; Gaps 7;

QY 90 EGYVQRTIFFKDDGNYKTRAEVKEPSDTLVNRIE-LKIDFKEDNLTGKIKFYNKSH 146
DB 632 KGYPEVALHFVQDE--KTPSLAECGNIETALEAKALCKNWKELGVALLOQGHQ 586
QY 149 NVYIMADKQKNGIKNF-----KIHUNTDGQVQ--LAHYQNTPTGQVFLVLPD 197
DE 689 IVEMQRTKNTDKVSLFLYLTGNTGKLEKIRMKKIATIRKMGCRYONALYAGS----- 241
QY 198 NHYLSTQSALSKDPNEKRDHMYLLGFVTAAGTTHGND 235
DB 742 ---VSEVRILKNGQYS- ---LAYLTAA--THGLDE 749

RESULT 5
IC SYLBAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucyl-tRNA synthetase) (Eukaryote)
GN LEUS OR H10921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID:727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN:RD / KW20 / ATCC 51907;
RX MEDLINE:95350630; PubMed:7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kervavage A.P., Rult J.J., Tomb J.E., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu S.-F., Glodde A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.E.,
RA Utterback T.R., Hanna M.C., Nguyen B.T., Saidek D.M., Brandon R.C.,
RA Fine L.D., Pritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Guelm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RA strain RD."
CC Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP -> L-leucine + tRNA(Leu) + AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-T AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL institution -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch)
CC
CC EMBL: J03774; AAC22581.1;
CC TIGR: H15521;
CC InterPro: IPR02302; Leu-tRNA-synth1a.
CC InterPro: IPR02303; tRNA-synt_1a.
CC InterPro: IPR01412; tRNA-synt_1.
CC Pfam: PF00133; tRNA-synt_1; 1.
CC PRINTS: PR00985; TRNASYNTHLEU.
CC TIGRFAMS: T-GR00396; leus_bact; 1.
CC PROSITE: PS00178; AA-TRNA-LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA: E59330AF5BAC6F7 CRC64;

Query Match 7.3%; Score 89.5; DB 1; Length 861;
Best Local Similarity 24.1%; Pred. No. 5.2;
Matches 45; Conservative 26; Mismatches 57; Indels 47; Gaps 9;

QY 49 TTGKLPVWPETLTILSYGQCFRYPDHMKRDFKFSAMPEGYVOERTIFKDDGNYKT 108
DB 314 TGKKLPI-WVANFVLMYHGGAVNVPADCRD--PEFAKYSLPKQVLAFLADEIDL 370
QY 109 RAEVFEQDTLVNRIE-LKIDFKEDNLTGKIKFYNKSHNYIMADK-QKNS-----KVN 164
DB 371 TKOAFVHGKLVNSDEDFGKNF--DGAENG-----IADKLEKLVGKQKVN 414
QY 165 FKIEH-----NIEGSGVLADHYQNTPTGSDPVLLENHYS-STQSAL 207
DB 415 YKCRDQVSKRYWGAPIMCTLENGDVPA-----PNEGLPIILPDVNVGKVRSP 467

QY 206 SKDPN 212
DB 465 NADPN 472

RESULT 6
IC HIS7_BSCA1 STANDARD; PRT; 353 AA.
AC P57203;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histidine biosynthesis bifunctional protein HisB (includes:
GN Hisidinol-phosphatase (EC 3.1.3.15); Imidazoleglycerol-phosphate
DE dehydratase (EC 4.2.1.19) (IGPD)).
GN HISB OR B0102.

OS *Bacillus subtilis* (subsp. *spizizenii*) (ATCC 6051) [Bacillus subtilis strain
 OS symbiotic bacterium].
 CC Bacteria: Proteobacteria: gamma subdivision: Bacillus.
 CX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 PC STRAIN=Tokyo 1598;
 RX MEDLINE=20445173; PubMed=10994077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.
 RI "Genome sequence of the endocellular bacterial symbiont of aphids
 CC *Bacillus* sp. AFS."
 CC Nature 407:81-86(2000).
 CC -!- CATALYTIC ACTIVITY: D-erythro-1-(imidazol-4-yl)glycerol 3-
 CC phosphate -> 3-(imidazol-4-yl)-2-oxopropyl phosphate + H(2O) - 1-histidinol
 CC + phosphate.
 CC -!- PATHWAY: Histidine biosynthesis: sixth step.
 CC -!- PATHWAY: Histidine biosynthesis: eighth step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HISTIDINOL-
 CC PHOSPHATASE FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE
 CC IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF061175; BAR: J28211;
 CC InterPro: IPR000807; IPR01;
 CC Pfam: PF06475; IPR01;
 CC ProDom: PD002282; IPR01;
 CC TIGRfam: TIGR01261; Euk_Norm: 1;
 CC PROSITE: PS00954; IPR_DEHYDRATASE_1; 1;
 CC PROSITE: PS00955; IPR_DEHYDRATASE_2; 1;
 CC Histidine biosynthesis: Multifunctional enzyme; Lyase; Hydrolase;
 CC Complete proteome.
 FT DOMAIN 1 164 HISTIDINOL-PHOSPHATASE
 FT DOMAIN 165 453 IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE
 SO SEQUENCE 453 AA: 40845 MW: 9048275.626 kDa
 Query Match 6.4% Score 84.00 E: 1.0e-154
 Best Local Similarity 24.0% Pred. No. 2.0
 Matches 49: Conservative 44: Mismatches 44: Gaps 10:
 CY 63 LSTGVQSPSRYPDPRKPKHDPKSNAPSPYQVE-----FKSNQNTKTRAVRP -- 121
 DB 111 LSTGVQSPSRYPDPRKPKHDPKSNAPSPYQVE-----FKSNQNTKTRAVRP -- 121
 CY 62 TSPFPLOPS- TALEFMISYERS---EGVFDGGLIGLHIFDQPTGKKIKRHLNML 126
 CY 115 -----EGDT- GNNRIKKIKDKEDGNLSHWRKNSNHNVMALQKN 153
 DB 116 ENLKKSVVGGSDMDLNNLKKIKKELCNKMLTETLKHNYTALDPK 174
 CY 160 GKVPFKPHNIE-----DGGVQLADHYQNTPEIGKPVLPKHYLET-SALKSTENK 214
 DB 175 ETKSVIKVWLDSELSKIDGKVFDMVLRGVSNGTGM- ---NLSVQGLDTHGHT 229
 CY 215 -SCHVWLLG--FVTAAGTIGHM 233
 DB 230 IEDTGIVLGEALLQALGKKNL 251
 RESULT 7
 D152_HAEIN
 ID D152_HAEIN STANDARD: PRI: 795 AA.
 AC P44935;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

OS Protective surface antigen D15 precursor (80 kDa D15 antigen)
 OS (D-15-Ag) (Outer membrane protein D15).
 CC Bacteria: Proteobacteria: gamma subdivision: Pasteurellales.
 CX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 PC STRAIN=Re / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Klevorace A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott C.D., Shirley R., Liu L.-F., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uiterback L.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Ouellet C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 RI H3";
 RI Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: J12773; AAC22575.1;
 CC InterPro: IPR00184; Bac_surfac_D15.
 CC Pfam: PF01103; Bac_Surface_Ag; 1;
 CC Antigen: Outer membrane; Signal: Complete proteome.
 FT SIGNAL 1 19 POTENTIAL
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15
 SO SEQUENCE 795 AA: 87478 MW: 885691FC22E6EDM4 CRG64;
 Query Match 6.4% Score 87.50 E: 1.0e-154
 Best Local Similarity 22.4% Pred. No. 7.1
 Matches 49: Conservative 44: Mismatches 73: Indels 63: Gaps 12:
 CY 64 LSTGVQSPSRYPDPRKPKHDPKSNAPSPYQVE-----FKSNQNTKTRAVRP 121
 DB 427 IGVQSPSRYPDPRKPKHDPKSNAPSPYQVE-----FKSNQNTKTRAVRP 121
 CY 99 FPKDGNVKTAEVKEGEGDANNRIKKIKDKEDGNLSHWRKNSNHNVMALQKN 147
 DB 99 FPKDGNVKTAEVKEGEGDANNRIKKIKDKEDGNLSHWRKNSNHNVMALQKN 147
 CY 487 FPKDGNVKTAEVKEGEGDANNRIKKIKDKEDGNLSHWRKNSNHNVMALQKN 542
 DB 487 FPKDGNVKTAEVKEGEGDANNRIKKIKDKEDGNLSHWRKNSNHNVMALQKN 542
 CY 148 HRYVIMADKQK-NGTKVN-----FKIRHN-----IEDGVQLADHYQNTPEIGKPVLPKHYLET-SALKSTENK 195
 DB 148 HRYVIMADKQK-NGTKVN-----FKIRHN-----IEDGVQLADHYQNTPEIGKPVLPKHYLET-SALKSTENK 195
 CY 544 ENLYIQSMKFKGNGIKTDFDSEFGNNVNSLNKRGVPTKGVKAS-----GG-GVVTI 593
 DB 544 ENLYIQSMKFKGNGIKTDFDSEFGNNVNSLNKRGVPTKGVKAS-----GG-GVVTI 593
 CY 196 P--DNHLYLTQSALSKDPNEKSHHVVLLSPVTAAGTIGHM 232
 DB 196 P--DNHLYLTQSALSKDPNEKSHHVVLLSPVTAAGTIGHM 232
 CY 544 PGSDNRYKXLSADVCGFYDLDKCHLWVWSAKASAGYANG 632
 DB 544 PGSDNRYKXLSADVCGFYDLDKCHLWVWSAKASAGYANG 632
 RESULT 8
 D152_HAEIN
 ID D152_HAEIN STANDARD: PRI: 797 AA.
 AC P46024;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DT Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DB (D-15-Ag) (Outer membrane protein D15).

OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Pasteurellaceae;
 CC Haemophilus
 CC NCBI_taxonomy:727;
 CC (11)
 RP SEQUENCE FROM N.A.
 RP STRAIN: Serotype B;
 EX MEDLINE:95255676; PubMed:7737523;
 RA Fick P.S., Loomore S., Chow P., Thomas W.B.,
 RT The sequencing of the 43-kDa D15 protective surface antigen of
 RT Haemophilus influenzae.";
 RT J. Gen. Microbiol. 136:97-99(1992).
 RN 121
 RP SEQUENCE FROM N.A.
 RP STRAIN: Mima and Egge / Serotype B;
 RX MEDLINE:97427952; PubMed:9284145;
 RA Loomore S.M., Yang Y.P., Coleman D.C., Shortland J.M., England D.M.,
 RA Klein M.H.;
 RT Outer membrane protein D-5 is conserved among Haemophilus influenzae
 RT species and may represent a universal protective antigen against
 RT invasive disease";
 RT Infect. Immun. 65:3701-3707(1997).
 CC 122
 CC 123
 CC 124
 CC 125
 CC 126
 CC 127
 CC 128
 CC 129
 CC 130
 CC 131
 CC 132
 CC 133
 CC 134
 CC 135
 CC 136
 CC 137
 CC 138
 CC 139
 CC 140
 CC 141
 CC 142
 CC 143
 CC 144
 CC 145
 CC 146
 CC 147
 CC 148
 CC 149
 CC 150
 CC 151
 CC 152
 CC 153
 CC 154
 CC 155
 CC 156
 CC 157
 CC 158
 CC 159
 CC 160
 CC 161
 CC 162
 CC 163
 CC 164
 CC 165
 CC 166
 CC 167
 CC 168
 CC 169
 CC 170
 CC 171
 CC 172
 CC 173
 CC 174
 CC 175
 CC 176
 CC 177
 CC 178
 CC 179
 CC 180
 CC 181
 CC 182
 CC 183
 CC 184
 CC 185
 CC 186
 CC 187
 CC 188
 CC 189
 CC 190
 CC 191
 CC 192
 CC 193
 CC 194
 CC 195
 CC 196
 CC 197
 CC 198
 CC 199
 CC 200
 CC 201
 CC 202
 CC 203
 CC 204
 CC 205
 CC 206
 CC 207
 CC 208
 CC 209
 CC 210
 CC 211
 CC 212
 CC 213
 CC 214
 CC 215
 CC 216
 CC 217
 CC 218
 CC 219
 CC 220
 CC 221
 CC 222
 CC 223
 CC 224
 CC 225
 CC 226
 CC 227
 CC 228
 CC 229
 CC 230
 CC 231
 CC 232
 CC 233
 CC 234
 CC 235
 CC 236
 CC 237
 CC 238
 CC 239
 CC 240
 CC 241
 CC 242
 CC 243
 CC 244
 CC 245
 CC 246
 CC 247
 CC 248
 CC 249
 CC 250
 CC 251
 CC 252
 CC 253
 CC 254
 CC 255
 CC 256
 CC 257
 CC 258
 CC 259
 CC 260
 CC 261
 CC 262
 CC 263
 CC 264
 CC 265
 CC 266
 CC 267
 CC 268
 CC 269
 CC 270
 CC 271
 CC 272
 CC 273
 CC 274
 CC 275
 CC 276
 CC 277
 CC 278
 CC 279
 CC 280
 CC 281
 CC 282
 CC 283
 CC 284
 CC 285
 CC 286
 CC 287
 CC 288
 CC 289
 CC 290
 CC 291
 CC 292
 CC 293
 CC 294
 CC 295
 CC 296
 CC 297
 CC 298
 CC 299
 CC 300
 CC 301
 CC 302
 CC 303
 CC 304
 CC 305
 CC 306
 CC 307
 CC 308
 CC 309
 CC 310
 CC 311
 CC 312
 CC 313
 CC 314
 CC 315
 CC 316
 CC 317
 CC 318
 CC 319
 CC 320
 CC 321
 CC 322
 CC 323
 CC 324
 CC 325
 CC 326
 CC 327
 CC 328
 CC 329
 CC 330
 CC 331
 CC 332
 CC 333
 CC 334
 CC 335
 CC 336
 CC 337
 CC 338
 CC 339
 CC 340
 CC 341
 CC 342
 CC 343
 CC 344
 CC 345
 CC 346
 CC 347
 CC 348
 CC 349
 CC 350
 CC 351
 CC 352
 CC 353
 CC 354
 CC 355
 CC 356
 CC 357
 CC 358
 CC 359
 CC 360
 CC 361
 CC 362
 CC 363
 CC 364
 CC 365
 CC 366
 CC 367
 CC 368
 CC 369
 CC 370
 CC 371
 CC 372
 CC 373
 CC 374
 CC 375
 CC 376
 CC 377
 CC 378
 CC 379
 CC 380
 CC 381
 CC 382
 CC 383
 CC 384
 CC 385
 CC 386
 CC 387
 CC 388
 CC 389
 CC 390
 CC 391
 CC 392
 CC 393
 CC 394
 CC 395
 CC 396
 CC 397
 CC 398
 CC 399
 CC 400
 CC 401
 CC 402
 CC 403
 CC 404
 CC 405
 CC 406
 CC 407
 CC 408
 CC 409
 CC 410
 CC 411
 CC 412
 CC 413
 CC 414
 CC 415
 CC 416
 CC 417
 CC 418
 CC 419
 CC 420
 CC 421
 CC 422
 CC 423
 CC 424
 CC 425
 CC 426
 CC 427
 CC 428
 CC 429
 CC 430
 CC 431
 CC 432
 CC 433
 CC 434
 CC 435
 CC 436
 CC 437
 CC 438
 CC 439
 CC 440
 CC 441
 CC 442
 CC 443
 CC 444
 CC 445
 CC 446
 CC 447
 CC 448
 CC 449
 CC 450
 CC 451
 CC 452
 CC 453
 CC 454
 CC 455
 CC 456
 CC 457
 CC 458
 CC 459
 CC 460
 CC 461
 CC 462
 CC 463
 CC 464
 CC 465
 CC 466
 CC 467
 CC 468
 CC 469
 CC 470
 CC 471
 CC 472
 CC 473
 CC 474
 CC 475
 CC 476
 CC 477
 CC 478
 CC 479
 CC 480
 CC 481
 CC 482
 CC 483
 CC 484
 CC 485
 CC 486
 CC 487
 CC 488
 CC 489
 CC 490
 CC 491
 CC 492
 CC 493
 CC 494
 CC 495
 CC 496
 CC 497
 CC 498
 CC 499
 CC 500
 CC 501
 CC 502
 CC 503
 CC 504
 CC 505
 CC 506
 CC 507
 CC 508
 CC 509
 CC 510
 CC 511
 CC 512
 CC 513
 CC 514
 CC 515
 CC 516
 CC 517
 CC 518
 CC 519
 CC 520
 CC 521
 CC 522
 CC 523
 CC 524
 CC 525
 CC 526
 CC 527
 CC 528
 CC 529
 CC 530
 CC 531
 CC 532
 CC 533
 CC 534
 CC 535
 CC 536
 CC 537
 CC 538
 CC 539
 CC 540
 CC 541
 CC 542
 CC 543
 CC 544
 CC 545
 CC 546
 CC 547
 CC 548
 CC 549
 CC 550
 CC 551
 CC 552
 CC 553
 CC 554
 CC 555
 CC 556
 CC 557
 CC 558
 CC 559
 CC 560
 CC 561
 CC 562
 CC 563
 CC 564
 CC 565
 CC 566
 CC 567
 CC 568
 CC 569
 CC 570
 CC 571
 CC 572
 CC 573
 CC 574
 CC 575
 CC 576
 CC 577
 CC 578
 CC 579
 CC 580
 CC 581
 CC 582
 CC 583
 CC 584
 CC 585
 CC 586
 CC 587
 CC 588
 CC 589
 CC 590
 CC 591
 CC 592
 CC 593
 CC 594
 CC 595
 CC 596
 CC 597
 CC 598
 CC 599
 CC 600
 CC 601
 CC 602
 CC 603
 CC 604
 CC 605
 CC 606
 CC 607
 CC 608
 CC 609
 CC 610
 CC 611
 CC 612
 CC 613
 CC 614
 CC 615
 CC 616
 CC 617
 CC 618
 CC 619
 CC 620
 CC 621
 CC 622
 CC 623
 CC 624
 CC 625
 CC 626
 CC 627
 CC 628
 CC 629
 CC 630
 CC 631
 CC 632
 CC 633
 CC 634
 CC 635
 CC 636
 CC 637
 CC 638
 CC 639
 CC 640
 CC 641
 CC 642
 CC 643
 CC 644
 CC 645
 CC 646
 CC 647
 CC 648
 CC 649
 CC 650
 CC 651
 CC 652
 CC 653
 CC 654
 CC 655
 CC 656
 CC 657
 CC 658
 CC 659
 CC 660
 CC 661
 CC 662
 CC 663
 CC 664
 CC 665
 CC 666
 CC 667
 CC 668
 CC 669
 CC 670
 CC 671
 CC 672
 CC 673
 CC 674
 CC 675
 CC 676
 CC 677
 CC 678
 CC 679
 CC 680
 CC 681
 CC 682
 CC 683
 CC 684
 CC 685
 CC 686
 CC 687
 CC 688
 CC 689
 CC 690
 CC 691
 CC 692
 CC 693
 CC 694
 CC 695
 CC 696
 CC 697
 CC 698
 CC 699
 CC 700
 CC 701
 CC 702
 CC 703
 CC 704
 CC 705
 CC 706
 CC 707
 CC 708
 CC 709
 CC 710
 CC 711
 CC 712
 CC 713
 CC 714
 CC 715
 CC 716
 CC 717
 CC 718
 CC 719
 CC 720
 CC 721
 CC 722
 CC 723
 CC 724
 CC 725
 CC 726
 CC 727
 CC 728
 CC 729
 CC 730
 CC 731
 CC 732
 CC 733
 CC 734
 CC 735
 CC 736
 CC 737
 CC 738
 CC 739
 CC 740
 CC 741
 CC 742
 CC 743
 CC 744
 CC 745
 CC 746
 CC 747
 CC 748
 CC 749
 CC 750
 CC 751
 CC 752
 CC 753
 CC 754
 CC 755
 CC 756
 CC 757
 CC 758
 CC 759
 CC 760
 CC 761
 CC 762
 CC 763
 CC 764
 CC 765
 CC 766
 CC 767
 CC 768
 CC 769
 CC 770
 CC 771
 CC 772
 CC 773
 CC 774
 CC 775
 CC 776
 CC 777
 CC 778
 CC 779
 CC 780
 CC 781
 CC 782
 CC 783
 CC 784
 CC 785
 CC 786
 CC 787
 CC 788
 CC 789
 CC 790
 CC 791
 CC 792
 CC 793
 CC 794
 CC 795
 CC 796
 CC 797
 CC 798
 CC 799
 CC 800
 CC 801
 CC 802
 CC 803
 CC 804
 CC 805
 CC 806
 CC 807
 CC 808
 CC 809
 CC 810
 CC 811
 CC 812
 CC 813
 CC 814
 CC 815
 CC 816
 CC 817
 CC 818
 CC 819
 CC 820
 CC 821
 CC 822
 CC 823
 CC 824
 CC 825
 CC 826
 CC 827
 CC 828
 CC 829
 CC 830
 CC 831
 CC 832
 CC 833
 CC 834
 CC 835
 CC 836
 CC 837
 CC 838
 CC 839
 CC 840
 CC 841
 CC 842
 CC 843
 CC 844
 CC 845
 CC 846
 CC 847
 CC 848
 CC 849
 CC 850
 CC 851
 CC 852
 CC 853
 CC 854
 CC 855
 CC 856
 CC 857
 CC 858
 CC 859
 CC 860
 CC 861
 CC 862
 CC 863
 CC 864
 CC 865
 CC 866
 CC 867
 CC 868
 CC 869
 CC 870
 CC 871
 CC 872
 CC 873
 CC 874
 CC 875
 CC 876
 CC 877
 CC 878
 CC 879
 CC 880
 CC 881
 CC 882
 CC 883
 CC 884
 CC 885
 CC 886
 CC 887
 CC 888
 CC 889
 CC 890
 CC 891
 CC 892
 CC 893
 CC 894
 CC 895
 CC 896
 CC 897
 CC 898
 CC 899
 CC 900
 CC 901
 CC 902
 CC 903
 CC 904
 CC 905
 CC 906
 CC 907
 CC 908
 CC 909
 CC 910
 CC 911
 CC 912
 CC 913
 CC 914
 CC 915
 CC 916
 CC 917
 CC 918
 CC 919
 CC 920
 CC 921
 CC 922
 CC 923
 CC 924
 CC 925
 CC 926
 CC 927
 CC 928
 CC 929
 CC 930
 CC 931
 CC 932
 CC 933
 CC 934
 CC 935
 CC 936
 CC 937
 CC 938
 CC 939
 CC 940
 CC 941
 CC 942
 CC 943
 CC 944
 CC 945
 CC 946
 CC 947
 CC 948
 CC 949
 CC 950
 CC 951
 CC 952
 CC 953
 CC 954
 CC 955
 CC 956
 CC 957
 CC 958
 CC 959
 CC 960
 CC 961
 CC 962
 CC 963
 CC 964
 CC 965
 CC 966
 CC 967
 CC 968
 CC 969
 CC 970
 CC 971
 CC 972
 CC 973
 CC 974
 CC 975
 CC 976
 CC 977
 CC 978
 CC 979
 CC 980
 CC 981
 CC 982
 CC 983
 CC 984
 CC 985
 CC 986
 CC 987
 CC 988
 CC 989
 CC 990
 CC 991
 CC 992
 CC 993
 CC 994
 CC 995
 CC 996
 CC 997
 CC 998
 CC 999
 CC 1000

GN COPA
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovine; Bos.
 CC NCBI_taxonomy:9913;
 CC (11)
 RP SEQUENCE FROM N.A.
 RP Strain F.1.1;
 RA Wickham F.1.1;
 RT FUNCTION: XENIN IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
 RT TO DILYSINE MOTIFS AND REVERSLY ASSOCIATES WITH GOLGI NON-
 RT GLATHIN-COATED VESICLES, WHICH FURTHER MEDIATE BIOSYNTHESIS
 RT OF PROTEIN TRANSPORT FROM THE ER VIA THE GOLGI UP TO THE TRANS GOLGI
 RT NETWORK. COMPLEX FORMATION IS REQUIRED FOR BUDDING FROM GOLGI
 RT MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GOLGI-TO-ER
 RT TRANSPORT OF DILYSINE-TAGGED PROTEINS. IN MAMMALS, THE COMPLEX
 RT CAN ONLY BE RECRUITED BY MEMBRANES ASSOCIATED TO ADP-RIBOSYLATION
 RT FACTORS (ARF5), WHICH ARE SMALL GTP-BINDING PROTEINS. THE COMPLEX
 RT ALSO INFLUENCES THE GOLGI STRUCTURAL INTEGRITY, AS WELL AS THE
 RT PROCESSING ACTIVITY, AND ENDOCYTIC RECYCLING OF LIL RECEPTORS (BY
 RT SIMILARITY).
 CC 121
 CC 122
 CC 123
 CC 124
 CC 125
 CC 126
 CC 127
 CC 128
 CC 129
 CC 130
 CC 131
 CC 132
 CC 133
 CC 134
 CC 135
 CC 136
 CC 137
 CC 138
 CC 139
 CC 140
 CC 141
 CC 142
 CC 143
 CC 144
 CC 145
 CC 146
 CC 147
 CC 148
 CC 149
 CC 150
 CC 151
 CC 152
 CC 153
 CC 154
 CC 155
 CC 156
 CC 157
 CC 158
 CC 159
 CC 160
 CC 161
 CC 162
 CC 163
 CC 164
 CC 165
 CC 166
 CC 167
 CC 168
 CC 169
 CC 170
 CC 171
 CC 172
 CC 173
 CC 174
 CC 175
 CC 176
 CC 177
 CC 178
 CC 179
 CC 180
 CC 181
 CC 182
 CC 183
 CC 184
 CC 185
 CC 186
 CC 187
 CC 188
 CC 189
 CC 190
 CC 191
 CC 192
 CC 193
 CC 194
 CC 195
 CC 196
 CC 197
 CC 198
 CC 199
 CC 200
 CC 201
 CC 202
 CC 203
 CC 204
 CC 205
 CC 206
 CC 207
 CC 208
 CC 209
 CC 210
 CC 211
 CC 212
 CC 213
 CC 214
 CC 215
 CC 216
 CC 217
 CC 218
 CC 219
 CC 220
 CC 221
 CC 222
 CC 223
 CC 224
 CC 225
 CC 226
 CC 227
 CC 228
 CC 229
 CC 230
 CC 231
 CC 232
 CC 233
 CC 234
 CC 235
 CC 236
 CC 237
 CC 238
 CC 239
 CC 240
 CC 241
 CC 242
 CC 243
 CC 244
 CC 245
 CC 246
 CC 247
 CC 248
 CC 249
 CC 250
 CC 251
 CC 252
 CC 253
 CC 254
 CC 255
 CC 256
 CC 257
 CC 258
 CC 259
 CC 260
 CC 261
 CC 262
 CC 263
 CC 264
 CC 265
 CC 266
 CC 267
 CC 268
 CC 269
 CC 270
 CC 271
 CC 272
 CC 273
 CC 274
 CC 275
 CC 276
 CC 277
 CC 278
 CC 279
 CC 280
 CC 281
 CC 282
 CC 283
 CC 284
 CC 285
 CC 286
 CC 287
 CC 288
 CC 289
 CC 290
 CC 291
 CC 292
 CC 293
 CC 294
 CC 295
 CC 296
 CC 297
 CC 298
 CC 299
 CC 300
 CC 301
 CC 302
 CC 303
 CC 304
 CC 305
 CC 306
 CC 307
 CC 308
 CC 309
 CC 310
 CC 311
 CC 312
 CC 313
 CC 314
 CC 315
 CC 316
 CC 317
 CC 318
 CC 319
 CC 320
 CC 321
 CC 322
 CC 323
 CC 324
 CC 325
 CC 326
 CC 327
 CC 328
 CC 329
 CC 330
 CC 331
 CC 332
 CC 333
 CC 334
 CC 335
 CC 336
 CC 337
 CC 338
 CC 339
 CC 340
 CC 341
 CC 342
 CC 343
 CC 344
 CC 345
 CC 346
 CC 347
 CC 348
 CC 349
 CC 350
 CC 351
 CC 352
 CC 353
 CC 354
 CC 355
 CC 356
 CC 357
 CC 358
 CC 359
 CC 360
 CC 361
 CC 362
 CC 363
 CC 364
 CC 365
 CC 366
 CC 367
 CC 368
 CC 369
 CC 370
 CC 371
 CC 372
 CC 373
 CC 374
 CC 375
 CC 376
 CC 377
 CC 378
 CC 379
 CC 380
 CC 381
 CC 382
 CC 383
 CC 384
 CC 385
 CC 386
 CC 387
 CC 388
 CC 389
 CC 390
 CC 391
 CC 392
 CC 393
 CC 394
 CC 395
 CC 396
 CC 397
 CC 398
 CC 399
 CC 400
 CC 401
 CC 402
 CC 403
 CC 404
 CC 405
 CC 406
 CC 407
 CC 408
 CC 409
 CC 410
 CC 411
 CC 412
 CC 413
 CC 414
 CC 415
 CC 416
 CC 417
 CC 418
 CC 419
 CC 420
 CC 421
 CC 422
 CC 423
 CC 424
 CC 425
 CC 426
 CC 427
 CC 428
 CC 429
 CC 430
 CC 431
 CC 432
 CC 433
 CC 434
 CC 435
 CC 436
 CC 437
 CC 438
 CC 439
 CC 440
 CC 441
 CC 442
 CC 443
 CC 444
 CC 445
 CC 446
 CC 447
 CC 448
 CC 449
 CC 450
 CC 451
 CC 452
 CC 453
 CC 454
 CC 455
 CC 456
 CC 457
 CC 458
 CC 459
 CC 460
 CC 461
 CC 462
 CC 463
 CC 464
 CC 465
 CC 466
 CC 467
 CC 468
 CC 469
 CC 470
 CC 471
 CC 472
 CC 473
 CC 474
 CC 475
 CC 476
 CC 477
 CC 478
 CC 479
 CC 480
 CC 481
 CC 482
 CC 483
 CC 484
 CC 485
 CC 486
 CC 487
 CC 488
 CC 489
 CC 490
 CC 491
 CC 492
 CC 493
 CC 494
 CC 495
 CC 496
 CC 497
 CC 498
 CC 499
 CC 500
 CC 501
 CC 502
 CC 503
 CC 504

```

100 609 DEVIHVRNKKAGGSLIAVAKKSTVEVALPPTVE--KTRSHALAEENINILIAA 447
101 126 KGIDFKEDGSLCHKLEKYNENNVYIMARKKNIRVND-- --K-KPENETENGV 170
102 666 KALDRKNCWKEGCEVALDQNGUJIVEMCVRIKKIKLEKSLVLTATNLEKIKAKKIKIAEL 145
103 277 C--LAHYQOONTEGCGVLLPNNHYLSGSLSKLKPENKRIHHVLAFLVLAAGLIHSDG 284
104 726 EKMSGHYUNALYGD-----VSEKVKIKRNGCKRS--LAVLAA--THGLD 769
105 235 E 235
106 269 E 269
107
108 RESULT 10
109 PROE_YEAST
110 ID PROE_YEAST STANDARDS: PRT: 2222 AA.
111 AC P21951:
112 DT 01-AUG-1991 (Rel. 19, Created)
113 DT 01-AUG-1991 (Rel. 19, Last sequence update)
114 DT 16-OCT-2002 (Rel. 43, Last annotation update)
115 DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
116 polymerase II subunit A).
117 GN POL2 OR DUN2 OR YNL262W OR NUG25.
118 OS Saccharomyces cerevisiae (Baker's yeast).
119 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
120 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces;
121 OX NBLT:taxID=4932;
122
123 (1)
124 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1214-1221.
125 RP MEDLINE 90381771; PubMed-2169342;
126 RA Sen-Gupta M., Lyck R., Selig U., Naedertal K.R., Hopmann F.H.,
127 Morrison A., Ataki H., Clark A.R., Hamatake R.K., Saito A.,
128 "A third essential DNA polymerase in S. cerevisiae".
129 PL Cell 62:1143-1151(1990);
130 [2]
131
132 RN SEQUENCE OF 1-221 FROM N.A.
133 RP STRAIN=S288c / FY1679;
134 RA Sen-Gupta M., Lyck R., Selig U., Naedertal K.R., Hopmann F.H.,
135 Morrison A., Ataki H., Clark A.R., Hamatake R.K., Saito A.,
136 "The sequence of a 24,152 bp segment from the left arm of chromosome
137 XIV from Saccharomyces cerevisiae between the PHU1 and the POL2
138 genes".
139 RL Yeast 12:505-514 (1995).
140 [3]
141
142 RN TEMPERATURE SENSITIVE MUTANTS.
143 RP MEDLINE-92164663; PubMed-1537345;
144 RA Ataki H., Ropp P.A., Johnson A.L., Johnston L.R., Morrison A.,
145 Saito A.,
146 "DNA polymerase II, the probable homolog of mammalian DNA polymerase
147 epsilon, replicates chromosomal DNA in the yeast Saccharomyces
148 cerevisiae".
149 RL FMO J. 11:733-740(1992).
150
151 (1)
152 RN FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
153 REPLICATION.
154 (2)
155 RN CATALYTIC ACTIVITY: N deoxyribonucleoside triphosphate - N diphosphate
156 + [DNA(N)].
157 (3)
158 RN SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 kDa, 130 kDa, 34 kDa, 40
159 kDa, AND 29 kDa).
160 (4)
161 RN SUBCELLULAR LOCATION: Nuclear.
162 (5)
163 RN DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
164 N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
165 FOR COMPLEXING SUBUNIT B AND C.
166 (6)
167 RN MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
168 ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
169 DIFFERENT REACTIONS OF DNA SYNTHESIS.
170 (7)
171 RN SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH
172 SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.
173
174 (8)
175 This SWISS-PROT entry is copyright. It is produced through a collaboration
176 between the Swiss Institute of Bioinformatics and the EMBL outstation -
177 the European Bioinformatics Institute. There are no restrictions on its

```

```

100 use of non-profit institutions as long as its content is in no way
101 modified and this statement is not removed. Usage by and for commercial
102 entities requires a license agreement (See http://www.isb.eth.ch/announce/
103 or send an email to license@isb.eth.ch).
104
105 DR EM5L: M60456; AAA69711.1;
106 DR EM5L: X92494; CAA63235.1;
107 DR EM5L: Z71538; CAA6169.1;
108 DR PIR: A36028; A36028;
109 DR SDB: SC005206; POL2.
110 DR InterPro: IPR002064; DNA_pol_B.
111 DR Pfam: PF01136; DNA_pol_B.1.
112 DR PRINTS: PR00106; DNAPOLB.
113 DR SMART: SM00466; POLB; 1.
114 DR PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
115 KW Transferase; DNA-directed DNA polymerase; DNA replication;
116 KW DNA-binding; Zinc-finger; Nuclear protein;
117 FT ZNFING 2108 2151 POTENTIAL.
118 FT VARIANT 644 644 M -> I (IN POL2-5 IS MUTANT).
119 FT VARIANT 710 710 P -> S (IN POL2-18 IS MUTANT).
120 SQ SEQUENCE 2222 AA; 25569 MW; CECDE2AB147D65B CRC64;
121
122 Query Match: 6.9%; Score 67.5; DB 1: Length 2222;
123 Best Local Similarity 28.2%; Pred. No. 24;
124 Matches 37; Conservative 14; Mismatches 49; Indels 31; Gaps 7;
125
126 QY 53 LPVWMP-TLVTTISGVGQPSKYPDH-----KRDFKSAPEGV---QERTI 98
127 DB 883 LPKSFETVFFLENGKKLYLSYPSKLNRYVHQFTNHCYQELKPLNYIETHSEMTI 942
128 QY 99 FFKDGGYKTR--AYKPGSDTLVNR-----IELKGIDFKEDNTGKLEKLYN 146
129 DB 943 FFEVDGPKAN-LPSSKEEGKIKKRYAVENEGSLALKGELKRGEL---QLIKNEQ 999
130 QY 147 S--HNVLMAD 155
131 DB 1500 SDIFKVFLEGE 1010
132
133 RESULT 11
134 TRNA_CAMJE
135 ID TRNA_CAMJE STANDARD: PRT: 357 AA.
136 AC QEP242:
137 DT 15-JUN-2002 (Rel. 41, Created);
138 DT 15-JUN-2002 (Rel. 41, Last sequence update);
139 DT 15-JUN-2002 (Rel. 41, Last annotation update);
140 DE TRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (CPNA(M-5-05))-
141 methyltransferase (MONT).
142 GN TRNA OF CJC810C.
143 OS Campylobacter jejuni.
144 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
145 OC Campylobacter.
146 OX NBLT:taxID=197;
147 [1]
148 RN SEQUENCE FROM K.A.
149 RP S-HAIN-MCIC 11165;
150 RA MEDLINE-20150912; PubMed-1038204;
151 RA Parkhill J., Wren B.W., Mangail K., Kelley J.M., Churcher C.,
152 Basset D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
153 Jorgensen K., Karyshev A.V., Moule S., Pallen M.J., Penn C.W.,
154 Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
155 Whitehead S., Barrall R.G.;
156 "The genome sequence of the food-borne pathogen Campylobacter jejuni
157 reveals hypervariable sequences".
158 RL Nature 431:665-668(2000).
159 (1)
160 RN FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
161 54 (M-5-U54) in all tRNA (By similarity).
162 (2)
163 RN CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
164 homocysteine + tRNA containing thymine.
165 (3)
166 RN SIMILARITY: BELONGS TO THE RNA M5C METHYLTRANSFERASE FAMILY. TRNA
167 SUBFAMILY.
168
169 This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL database
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk.)

CC EMBL: AL39076; CAB73026.1; -
 DR InterPro: IPR000091; SAM_Pding.
 DR InterPro: IPR001566; TRMA.
 DR PROSITE: PS01233; TRMA.1; 1.
 DR PROSITE: PS01231; TRMA.2; FALSE_NEG.
 KW Transferrase; Methyltransferase; RNA processing; Comp. and molec.
 FT DOMAIN 207 213 S-ADENOSYL-METHIONINE BINDING (RY
 FT ACT_SITE 415 415 RY SIMILARITY)
 FT ACT_SITE 415 415 RY SIMILARITY
 SQ SEQUENCE 357 AA: 42276 MW: 653328347.67 ER497 CR044.

Query Match 6.84; Score 87; DB 1; Length 417;
 Best Local Similarity 24.84; Pred. No. 1.9;
 Matches 30; Conservativeness 18; Mismatches 39; Indels 34; Gaps 50

QY 79 KRHPFKSAMPEYVORLFFFDGQNYKTRAVKP-FOGIVY-----NATLELS-227
 DB 14 EKRSFKKYEKFTKDKLFASKKHVTRAEISFYHNMFFVAMFTKSEKTYLIV-73
 QY 126 IDKED-----DNTHGKSYNNYSHVYMAKKNKJFKYKTRKINIP-172
 DB 74 LDFADKLDAMFPLLYLQDKNKKK-----GVEVETFKDQ-LSLITLHKH-125
 QY 173 D 173
 DB 126 D 126

RESULT 12
 SV62_DISCM
 ID SV62_DISCM STANDARD: PRT: 415 AA.
 AC P24506;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DR Synaptotagmin B (Synaptic vesicle protein G-p45-B).
 GN P65-B.
 OS Discosoma ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hymenoptera; Batoidae;
 OC Torpediniformes; Narcinidae; Torpedinidae; Discosomidae;
 OC NCHI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=91273951; PubMed=2554189;
 RA Wendland B., Miller K.S., Schalling G., Scheller K.H.;
 RT "Differential expression of the p65 gene family";
 RL Neuron 6:993-1007(1991).
 CC -!- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE MEMBRANE INTERACTIONS
 CC DURING TRAFFICKING OF SYNAPTIC VESICLES AT THE ACTIVE ZONE OF THE
 CC SYNAPSE. IT BINDS ACIDIC PHOSPHOLIPIDS WITH A SPECIFICITY THAT
 CC REQUIRES THE PRESENCE OF BOTH AN ACIDIC HEAD GROUP AND A GLYCYL
 CC BACKBONE.
 CC -!- SUBUNIT: HOMODIMER OR HOMOTRIMER (POSSIBLE).
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLES IN NEURONS.
 CC -!- TISSUE SPECIFICITY: SPINAL CORD, BRAINSTEM, MIDBRAIN AND ELVENTRIC
 CC ORGAN.
 CC -!- SIMILARITY: CONTAINS 2 C2 DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk.)

CC or send an email to license@ebi.ac.uk.)

CC EMBL: M64276; AAA49228.1; -
 DR PIR: JH0414; JH0414.
 DR HSSP: P21707; IRYN.
 DR InterPro: IPR000008; C2.
 DR InterPro: IPR002149; Lvl.
 DR InterPro: IPR001565; Synaptotagmin.
 DR Pfam: PF00148; C2; 2.
 DR PRINTS: PR00360; C2DOMAIN.
 DR PRINTS: PR00399; SYNAPTOTAGMIN.
 DR SMART: SM00239; C2; 2.
 DR PROSITE: PS00499; C2 DOMAIN 1; 2.
 DR PROSITE: PS00004; C2 DOMAIN 2; 2.
 KW Transmembrane; Repeat; Synapse; Multigene family; Glycoprotein.
 FT DOMAIN 1 74 VESICULAR (POTENTIAL).
 FT TRANSMEM 75 101 POTENTIAL.
 FT DOMAIN 102 435 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 153 399 PHOSPHOLIPID BINDING (HOMOBABE).
 FT DOMAIN 173 262 C2 DOMAIN 1.
 FT DOMAIN 304 395 C2 DOMAIN 2.
 FT CARBOHYD 46 6 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 439 AA: 44278 MW: 203395FD9C9F39 CR064.

Query Match 6.84; Score 87; DB 1; Length 439;
 Best Local Similarity 26.14; Pred. No. 1.8;
 Matches 58; Conservativeness 44; Mismatches 92; Indels 34; Gaps 13;

QY 14 SVELLGQVNGHKNVSGRESDATYKLTLPFTTKLAV-PWPTVYTLISLV-58
 DB 14 MMEIDTGNSTAGVFGSKNE-VFEK-KEKEMFLOKIPPLPFWALALATVSSILLETG 96
 QY 53 -----QFSSAYPDHMKRDEKFSKMPGYYQVETIFPKDGNKYKTRKSKRP----- 116
 DB 37 CLCICKKCKCKKKKKKKKKKK-----KNDINMK-----EWAGSGGNDDD 138
 QY 117 --DTLVNRIELKGDIFKRDGNI--LGHKIRYNYKSH----- 148
 DB 139 DAETQLTEGDEKEEKEEKLKGTQFSLDYFOANQLTVGLCAELPALQWGSTDPY 198
 QY 149 -NVTYNAQKN-KIKVN-----FKLHNFTDGGVOLA-----SHYQNTP 187
 DB 199 VKVFLPDKKKKYETKVKTKLTNPFTNESFVKVPYQ-ELCKKTMMAVYDFDKSKHC 257
 QY 188 ICGGVPVLLPD-----NHVLSGALSCKDKPKRDHNVILGHPVTAAG 228
 DB 258 IGVTVLTKYDVGGOGLREWRDLSEAESEKELCDICTSLNYVYVTAAG 305

RESULT 13
 D151_EAEIN
 ID D151_EAEIN STANDARD: PRT: 793 AA.
 AC Q32629;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D-15 Ag) (outer membrane protein D15).
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 OC Haemophilus.
 OC NCHI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PAK 12085;
 RA MEDLINE=97427952; PubMed=9264140;
 RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
 RA Kleit M.H.;
 RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
 RT species and may represent a universal protective antigen against
 RT invasive disease";
 KI Infect. Immun. 65:3701-3707(1997).
 RN [1]

CC -- SUBCELLULAR LOCATION: Outer membrane.
 CC -- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U60834; AAR61977.1; ..
 DR InterPro: IPR000194; BacSurfA.D5.
 DR Pfam: PF01103; BacSurfA.CA.1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 19 .. POTENTIAL.
 FT CHAIN 20 793 .. PROTECTIVE SURFACE ANTIGEN D15.
 SQ SEQUENCE 793 AA: 8751; MW: 5136203680; A14 CR654;
 Query Match 6.9%; Score 86.5; DB 1; Length 793;
 Best Local Similarity 22.9%; Pred. No. 8.6;
 Matches 50; Conservative 32; Mismatches 74; Indels 53; Gaps 12;
 QY 64 LFGVOCFSRYDHMKRHPF-----FKSMFEGVDS-----FTI 98
 DB 427 IGVTFSGSYQTSIKQDNFLGTGAASANTKNTYIGSVNIQYFPTKPGVSDAGNI 486
 QY 99 FPKDDNNTVTRAEVKEPQDTLVNRIELKQDEKFGNI--IGH-----KLPYRNS 167
 DB 487 FPKYNSKSDSYSSNYKRYTASNVTL-GPPYNNNSYVGLHYNYKLSNFAEYN-- 542
 QY 148 HNYVTADKQK-NGSKVN-----FRIRHN-----IFHGVGVGADRYGVNTHLIGGVVGL 195
 DB 543 KNAVOSMKFKNGIKINIDFESPKWYNS-KRGVTF-KGVKAS-------GSGGVTF 543
 QY 195 P--HNYLSTOSALSQKPKMKHMVLLGVTAALTHG 232
 DB 594 PGSDNKKYKLSADVOYFVLDQDHVWVSASAGYANG 642
 RESULT 14
 CAN2_HUMAN STANDARD: PRT: 700 AA.
 AC E17655; G16738.
 DT 01-AGU-1995 (Ref. 15, Created)
 DT 01-AGU-1990 (Ref. 15, Last sequence update)
 DT 15-JUN-2002 (Ref. 41, Last annotation update)
 DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
 DE (Calcium-activated neutral protease) (CAN) (M-type) (Microcalpain)
 DE (Molecular-calpain).
 GN CAPN2 OR CAPNL2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxID:9606;
 [1]
 SP SEQUENCE FROM N.A.
 RP MEDLINE:89166474; PubMed:2852352;
 RA Imajoh S., Aoki K., Ohno S., Etori Y., Kawasaki H., Saitoh H.,
 RA Suzuki K.;
 RT "Molecular cloning of the cDNA for the large subunit of the
 RT high-Ca2+-requiring form of human Ca2+-activated neutral protease."
 RL Biochemistry 27:8122-8126(1988).
 RN [2]
 RP SEQUENCE OF 1-75 FROM N.A.
 RC Tissue-Lymph node;
 RX MEDLINE:95197547; PubMed:25939481;
 RA Hata A., Ohno S., Akita Y., Suzuki K.;
 RT "Tandemly reiterated negative enhancer-like elements regulate
 RT transcription of a human gene for the large subunit of calcium-
 RT dependent protease."
 RL J. Biol. Chem. 264:6404-6411(1989).
 RN [3]

X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE:20105516; PubMed:12630123;
 RA Strobl S., Fernandez-Catalan C., Brann M., Huber R., Masamoto H.,
 RA Nakagawa K., Irie A., Sorinachi H., Bourenkow G., Bartunik H.,
 RA Suzuki K., Bode W.;
 RT "The crystal structure of calcium-free human calpain suggests an
 RT electrostatic switch mechanism for activation by calcium."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:568-572(2000).
 CC -- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with leu or val as the P2 residue.
 CC -- COFACTOR: Binds 3 calcium ions.
 CC -- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding.
 CC -- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: M4254; AAA35645.1; ..
 DR EMBL: J04700; AAA52760.1; ..
 DR PIR: A34218; A34218.
 DR PUB: 1KFO; 07-DEC-01.
 DR MEROPS: C02.C02; ..
 DR GeneID: 642479; CAPN2.
 DR MIM: 114230; ..
 DR InterPro: IPR022046; EF-hand.
 DR InterPro: IPR01300; Protease_C2.
 DR InterPro: IPR00169; SHprot_apsite.
 DR Pfam: PF00036; efband; 3.
 DR Pfam: PF00645; Peptidase_C2; 1.
 DR Pfam: PF01067; Calpain_III; 1.
 DR PRISM: PR03704; CALPAIN.
 DR PRODOM: P000012; EF-hand; 1.
 DR SMART: SM00200; Cyspep; 1.
 DR SMART: SM0054; Efh; 2.
 DR PROSITE: PS00018; EF-HAND; 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE-NEG.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE-NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Repeat; Mollusce family;
 KW 3D-structure.
 FT PROPEP 1 19 .. ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
 FT CHAIN 20 700 .. CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
 FT DOMAIN 20 210 .. THIOL PROTEASE DOMAIN 1.
 FT DOMAIN 21 355 .. THIOL PROTEASE DOMAIN 2.
 FT DOMAIN 356 514 .. DOMAIN 11, C2-LIKE DOMAIN.
 FT DOMAIN 515 529 .. LINKER.
 FT DOMAIN 530 700 .. DOMAIN IV.
 FT CA_BIND 541 552 .. EF-HAND 1.
 FT CA_BIND 585 596 .. EF-HAND 2.
 FT CA_BIND 615 626 .. EF-HAND 3.
 FT DOMAIN 650 661 .. ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 680 691 .. ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 105 125 .. BY SIMILARITY.
 FT ACT_SITE 262 262 .. BY SIMILARITY.
 FT ACT_SITE 286 286 .. BY SIMILARITY.
 FT CONFLICT 68 68 .. S -> G (IN REF. 2).
 FT CONFLICT 73 74 .. MR -> IE (IN REF. 2).
 SQ SEQUENCE 700 AA: 80066 MW: A944D13BC846531 CRC64;

GenCore version: 5.1.6
Copyright (c) 1993 - 2003 CompuLink Ltd.

OK protein - protein search, using sw model

Run on: July 16, 2003, 16:04:17 : Search time 15 seconds
(without alignments)
466,444 Million cells processed/sec

Title: US-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKCEELFTGVVPLVSLDQ3.....VLLFFVTAHHTHRELK 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2943222 residues

Total number of hits satisfying chosen parameters: 2/274

Minimum DB seq length: 0

Maximum DB seq length: 240000000

Post-processing: Minimum Match 9%

Listing first 45 summaries

Database: Issued_Ratons_AA*

- 1: /seq2.6/prodata/1/1aa/5A_5MB.pept
- 2: /seq2.6/prodata/1/1aa/5P_5MB.pept
- 3: /seq2.6/prodata/1/1aa/6A_5MB.pept
- 4: /seq2.6/prodata/1/1aa/6B_5MB.pept
- 5: /seq2.6/prodata/1/1aa/6C_5MB.pept
- 6: /seq2.6/prodata/1/1aa/6D_5MB.pept

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1266	99.2	236	3	US-08-893-327-16
2	1266	99.2	247	3	US-08-893-327-16
3	1262	98.9	236	1	US-08-753-143-2
4	1262	98.9	238	2	US-08-674-355-2
5	1262	98.9	238	2	US-08-680-376-2
6	1262	98.9	238	2	US-08-792-553-2
7	1262	98.9	238	3	US-08-753-144-2
8	1262	98.9	238	4	US-09-004-359-2
9	1262	98.9	238	4	US-09-172-063-2
10	1262	98.9	238	4	US-09-263-935-2
11	1262	98.9	236	4	US-08-727-452-2
12	1262	98.9	236	4	US-09-419-785-1
13	1262	98.9	1070	4	US-09-091-012A-2
14	1262	98.9	1552	4	US-09-127-237-2
15	1266	98.6	234	1	US-06-737-915A-2
16	1266	98.6	234	1	US-09-121-516-1
17	1266	98.6	236	4	US-08-140-072-2
18	1257	98.5	249	3	US-08-643-538-2
19	1257	98.5	249	4	US-09-503-222-2
20	1252	98.1	236	3	US-08-643-704A-45
21	1251	98.0	239	4	US-09-172-064-3
22	1251	98.0	239	4	US-09-513-783A-46
23	1251	98.0	239	4	US-09-318-925-4
24	1251	98.0	281	4	US-09-062-102-1
25	1251	98.0	281	4	US-09-354-945-1
26	1251	98.0	294	4	US-09-513-783A-2
27	1251	98.0	323	4	US-09-172-063-2

28	1251	98.0	364	4	US-09-085-305-6	Sequence 3, Appli
29	1251	98.0	459	4	US-09-513-783A-170	Sequence 170, App
30	1251	98.0	642	2	US-08-818-253-2	Sequence 2, Appli
31	1251	98.0	642	2	US-08-818-253-6	Sequence 6, Appli
32	1251	98.0	642	4	US-08-818-252-2	Sequence 2, Appli
33	1251	98.0	642	4	US-08-818-252-6	Sequence 6, Appli
34	1251	98.0	652	2	US-08-818-253-4	Sequence 4, Appli
35	1251	98.0	652	4	US-08-818-252-4	Sequence 4, Appli
36	1251	98.0	753	4	US-09-513-783A-176	Sequence 176, App
37	1251	98.0	815	4	US-09-513-783A-178	Sequence 178, App
38	1251	98.0	890	4	US-09-513-783A-174	Sequence 174, App
39	1251	98.0	941	4	US-09-513-783A-172	Sequence 172, App
40	1251	98.0	1407	4	US-06-974-543A-628	Sequence 628, App
41	1249	97.9	238	2	US-08-588-201-2	Sequence 2, Appli
42	1249	97.9	238	2	US-09-169-605-2	Sequence 2, Appli
43	1249	97.9	238	3	US-08-893-327-2	Sequence 2, Appli
44	1249	97.9	239	4	US-09-513-783A-52	Sequence 52, Appli
45	1249	97.9	247	3	US-08-893-327-20	Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-08-893-327-16
Sequence 16, Application US/08d93327
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Hauswirth, William W.
APPLICANT: Muzyczka, Nicholas
TITLE OF INVENTION: Humanized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arcold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,327
FILING DATE:
CLASSIFICATION: 5.4
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 9P/588,231
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: CFA:062\KIT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-893-327-16

Query Match: 99.2% Score 1246; DB 3; Length 236;
Best Local Similarity: 99.2%; Pred. No. 1.6e-130;
Matches 236; Conservative 1; Mismatches 1; Indels 0;
Gaps 0;
QY 1 MSKCEELFTGVVPLVSLDQ3VSGGSGESDATYGNILKLFICTGKLPVWPTL 60

Db 1 MSKGEELTGVPVILVELDGVNGHKFSVSSEGGDATYGGKLTGKLPCTTGKLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 120
Db 61 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 120
Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238

RESULT 2
US-08-893-327-18
Sequence 18, Application US/06893327
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Hauswirth, William W.
APPLICANT: Mazycka, Nicholas
TITLE OF INVENTION: Humanized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,327
FILING DATE:
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US 08/588,201
FILING DATE: 18-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,328
REFERENCE/DOCKET NUMBER: UPLA-062\NKT
TELEPHONE: (512) 418-3000
TELEFAX: (713) 743-3679
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-893-327-18

Query Match 99.2% Score 1266; DB 3; Length 247;
Best Local Similarity 99.2% Pred. No. 1.76 130;
Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MSKGEELTGVPVILVELDGVNGHKFSVSSEGGDATYGGKLTGKLPCTTGKLVFWPTL 60
Db 10 MSKGEELTGVPVILVELDGVNGHKFSVSSEGGDATYGGKLTGKLPCTTGKLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 120
Db 70 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 129

Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238

RESULT 3
US-08-753-143-2
Sequence 2, Application US/08753143A
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Heim, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
FILE REFERENCE: 07257/012003
CURRENT APPLICATION NUMBER: US/08/753,143A
CURRENT FILING DATE: 1996-11-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-08-753-143-2

Query Match 98.9% Score 1262; DB 1; Length 238;
Best Local Similarity 98.7% Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MSKGEELTGVPVILVELDGVNGHKFSVSSEGGDATYGGKLTGKLPCTTGKLVFWPTL 60
Db 1 MSKGEELTGVPVILVELDGVNGHKFSVSSEGGDATYGGKLTGKLPCTTGKLVFWPTL 60
Qy 61 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 120
Db 61 VTTLTYGQCFSRYPDHKKRHDFKFSAMPEGYOERTIFFKDGNYKTRAEVKEGDLV 120
Qy 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Db 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKKNIGKYNFKIRHNTEGSGVOLAD 180
Qy 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHVLSIQSALSQKDPNEKRDHNVLEFVTAAGITRGMDELYK 238

RESULT 4
US-08-679-865 2
Sequence 2, Application US/0879865
Patent No. 5912137
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
APPLICANT: Cubit, Andrew R.
TITLE OF INVENTION: Assays for Protein Kinases Using
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/679,865
FILING DATE: 16-JUL-1996

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sterella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 02-072-069200
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

Query Match          98.9%   Score 1262; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60
DB 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60
QY 61 VTLLYGVOCFSRYPOHMKRHDFEFSKAMPEGVVQERTIFFKEDSNYKTRAEVKFESGLV 120
DB 61 VTLLYGVOCFSRYPOHMKRHDFEFSKAMPEGVVQERTIFFKEDSNYKTRAEVKFESGLV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNNSHNRYIMALKQKNGIKVNEKIKENFTGGVGLAD 180
DB 121 NRLEKLGIDFKEDGNLGHKLEYNNNSHNRYIMALKQKNGIKVNEKIKENFTGGVGLAD 180
QY 181 HYQONTPIGDGVPLPNNHYLSTQSAISKDPNEKKDHVLLFEFVTAAGTTHGMJELK 238
DB 181 HYQONTPIGDGVPLPNNHYLSTQSAISKDPNEKKDHVLLFEFVTAAGTTHGMJELK 238

RESULT 5
US-08-680-876-2
Sequence 2, Application: US/08-680876
Patent No. 592558
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: Assays for Protein Kinases Using
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,876
FILING DATE: 16-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sterella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 02-072-069200
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

```

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-680-876-2

Query Match          98.9%   Score 1262; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60
DB 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60
QY 61 VTLLYGVOCFSRYPOHMKRHDFEFSKAMPEGVVQERTIFFKEDSNYKTRAEVKFESGLV 120
DB 61 VTLLYGVOCFSRYPOHMKRHDFEFSKAMPEGVVQERTIFFKEDSNYKTRAEVKFESGLV 120
QY 121 NRLEKLGIDFKEDGNLGHKLEYNNNSHNRYIMALKQKNGIKVNEKIKENFTGGVGLAD 180
DB 121 NRLEKLGIDFKEDGNLGHKLEYNNNSHNRYIMALKQKNGIKVNEKIKENFTGGVGLAD 180
QY 181 HYQONTPIGDGVPLPNNHYLSTQSAISKDPNEKKDHVLLFEFVTAAGTTHGMJELK 238
DB 181 HYQONTPIGDGVPLPNNHYLSTQSAISKDPNEKKDHVLLFEFVTAAGTTHGMJELK 238

RESULT 6
US-08-692-553-2
Sequence 2, Application: US/98792553
Patent No. 5981203
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: FISH & RICHARDSON P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/792,553
FILING DATE: 31-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lisa A. Hallie, Ph.D.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/941001/DC 96-160-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5059
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-692-553-2

Query Match          98.9%   Score 1252; DB 2; Length 238;
Best Local Similarity 98.7%   Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60
DB 1 MSKGEELFTGVVPIVLVELGDVNGVNHKFSVSGEGEDATYGLTKLTKPCTTCKLAVPWPPTL 60

```

```

QY 61 VTTLTYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120
DB 61 VTTSYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 120
DB 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 120
QY 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238
DB 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238

RESULT 7
US-08-753-144-2
: Sequence 2, Application US/08753144
: Patent No. 6066476
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Holm, Roger
: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/753,144
: FILING DATE: 20-NOV-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/727,452
: FILING DATE: 10-OCT-1996
: APPLICATION NUMBER: US95/14692
: FILING DATE: 13-NOV-1995
: APPLICATION NUMBER: 08/397,915
: FILING DATE: 13-NOV-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Halper, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07257/032002
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: Protein
: FRAGMENT TYPE: Internal
US-08-753-144-2

```

```

Query Match 98.98; Score 1262; DS 4; Length 238;
Best Local Similarity 98.78; Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVELDGVNNGHKFSVSGEGSDATYKLTKEICTGKLPVNPVL 60
DB 1 MSKGEELFTGVVPLVELDGVNNGHKFSVSGEGSDATYKLTKEICTGKLPVNPVL 60
QY 61 VTTLTYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120
DB 61 VTTSYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120

```

```

QY 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 180
DB 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 180
QY 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238
DB 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238

RESULT 8
US-09-094-359-2
: Sequence 2, Application US/09094359
: Patent No. 6143132
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Miyawaki, Atsushi
: APPLICANT: Llopis, Juan
: TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
: MEASURING THE PH OF A BIOLOGICAL SAMPLE
: FILE REFERENCE: 07257/067001
: CURRENT APPLICATION NUMBER: US/09/094,359
: CURRENT FILING DATE: 1998-06-09
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 238
: TYPE: PRT
: ORGANISM: Aequorea victoria
US-09-094-359-2

```

```

Query Match 98.98; Score 1262; DS 4; Length 238;
Best Local Similarity 98.78; Pred. No. 4.4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVELDGVNNGHKFSVSGEGSDATYKLTKEICTGKLPVNPVL 60
DB 1 MSKGEELFTGVVPLVELDGVNNGHKFSVSGEGSDATYKLTKEICTGKLPVNPVL 60
QY 61 VTTLTYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120
DB 61 VTTSYGVOCESRYPDHMKRHDFFKSAPEGVQERTFFKCDGNKFKVNEKVEGVLAD 120
QY 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 180
DB 121 NRLEKGIQDFKEDGNILGHKLEYNYSNHNVTIMADKQKNGIKVNEKVEGVLAD 180
QY 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238
DB 181 HYQONTPIGGPVLDPKHNYLSTOSALKSKDPKPKKHVLLFEVTAAGITDGMELYLK 238

```

```

RESULT 9
US-09-172-063-2
: Sequence 2, Application US/09172063
: Patent No. 6150176
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Miyawaki, Atsushi
: APPLICANT: Llopis, Juan
: APPLICANT: Wichter, Karelka M.
: APPLICANT: Remington, S. James
: TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
: MEASURING THE PH OF A BIOLOGICAL SAMPLE
: FILE REFERENCE: 07257/071001
: CURRENT APPLICATION NUMBER: US/09/172,063
: CURRENT FILING DATE: 1998-10-13
: EARLIER APPLICATION NUMBER: 09/094,359
: EARLIER FILING DATE: 1998-06-09
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 238
: TYPE: PRT

```

```

1  MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTGKLTCTTKLPPWPTL 60
DB 1  MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTGKLTCTTKLPPWPTL 60

QY 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180
DB 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180

QY 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180
DB 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180

QY 181  HYQONTPIGDGPVLLPDHNYLSTQSALSQKDPNEKRDHVVLEFVTAAGITHKMDLYK 238
DB 181  HYQONTPIGDGPVLLPDHNYLSTQSALSQKDPNEKRDHVVLEFVTAAGITHKMDLYK 238

RESULT 11
US-09-727-452-2
: Sequence 2, Application US/08727452A
: Patent No. 6315669
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Tsien, Roger Y.
: TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES
: FILE REFERENCE: 07257/032001
: CURRENT APPLICATION NUMBER: US/08/727,452A
: CURRENT FILING DATE: 1996-03-26
: EARLIER APPLICATION NUMBER: INT/US95/14692
: EARLIER FILING DATE: 1995-11-13
: EARLIER APPLICATION NUMBER: US 07/337,415
: EARLIER FILING DATE: 1994-11-10
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 2
: LENGTH: 248
: TYPE: PRT
: ORGANISM: Aequorea victoria
US-09-727-452-2

Query Match 95.9%; Score 1262; ID 4; Length 238;
Best Local Similarity 98.7%; Pred. No. 4, 4e-130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1  MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTGKLTCTTKLPPWPTL 60
DB 1  MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYGKLTGKLTCTTKLPPWPTL 60

QY 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 61  VTTLYGVQCSRYPDHMKRHDFKSAHPGEGVQVQRTIFFKDDGNYKTRAEVKFEGDTLV 120

QY 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180
DB 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180

QY 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180
DB 121  NRIELKGIQDFKEDGNLGHKLEYNSNSHNYIMAKQKNGIKVNFKIRHNTEGGVGLAD 180

QY 181  HYQONTPIGDGPVLLPDHNYLSTQSALSQKDPNEKRDHVVLEFVTAAGITHKMDLYK 238
DB 181  HYQONTPIGDGPVLLPDHNYLSTQSALSQKDPNEKRDHVVLEFVTAAGITHKMDLYK 238

RESULT 12
US-09-418-725-2
: Sequence 1, Application US/09418785
: Patent No. 6414119
: GENERAL INFORMATION:
: APPLICANT: Fisher, Hugh
: APPLICANT: Fisher, Hugh
: TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
: FILE REFERENCE: the Aequorea victoria Green Fluorescent Protein
: CURRENT APPLICATION NUMBER: US/09/418,785
: CURRENT FILING DATE: 1999-10-15
: PRIOR APPLICATION NUMBER: 60/104,563
: PRIOR FILING DATE: 1998-10-16

```

```

1  NUMBER OF SEQ ID NOS: 3
2  SOFTWARE: FastSeq for Windows Version 4.0
3  SEQ ID NO 1
4  LENGTH: 238
5  TYPE: PRT
6  ORGANISM: Aquorea victoria
7  PUBLICATION INFORMATION:
8  AUTHORS: Prasher, D.C. et al.
9  TITLE: Primary structure of the Aquorea victoria green-
10  JOURNAL: Gene
11  VOLUME: 121
12  PAGES: 229-233
13  DATE: 1992-01-01
14  DATABASE ACCESSION NUMBER: GenBank M6255
15  DATABASE ENTRY DATE: 1993-04-26
16  US-09-418-785-1

Query Match      98.9%  Score 1262; DB 4; Length 238;
Best Local Similarity 98.7%  Pred. No. 4 to 130;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKCTTCKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKCTTCKLPVWPPTL 60
QY 61 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 120
DB 61 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 120
QY 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 180
DB 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 180
QY 181 HYQONTPIGDGPVLLPDNNHYLSLTSALSKDNEKRDHNVLEFVTAAGITGHMDELYK 238
DB 181 HYQONTPIGDGPVLLPDNNHYLSLTSALSKDNEKRDHNVLEFVTAAGITGHMDELYK 238

```

```

1  REGISTRATION NUMBER: 38295
2  REFERENCE/PACKET NUMBER: 14014.9183
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: 404-688-6770
5  TELEFAX: 404-688-9880
6  INFORMATION FOR SEQ ID NO: 2:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 1070 amino acids
9  TYPE: amino acid
10  TOPOLOGY: linear
11  MOLECULE TYPE: protein
12  SEQUENCE DESCRIPTION: SEQ ID NO: 2:
13  US-09-091-042A-2

Query Match      98.9%  Score 1262; DB 4; Length 1070;
Best Local Similarity 98.7%  Pred. No. 4.2e+129;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKICTTCKLPVWPPTL 60
DB 23 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKICTTCKLPVWPPTL 82
QY 61 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 120
DB 83 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 142
QY 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 180
DB 143 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 202
QY 181 HYQONTPIGDGPVLLPDNNHYLSLTSALSKDNEKRDHNVLEFVTAAGITGHMDELYK 238
DB 203 HYQONTPIGDGPVLLPDNNHYLSLTSALSKDNEKRDHNVLEFVTAAGITGHMDELYK 260

RESULT 14
US-09-127-227-2
Sequence 2, Application US/09:27227
Patent No. 6393354
GENERAL INFORMATION:
APPLICANT: David M. Kolpe
APPLICANT: Travis J. Taylor
APPLICANT: Elizabeth McNamee
TITLE OF INVENTION: Replication-Competent Virus Expressing A
FILE REFERENCE: H098-05
CURRENT APPLICATION NUMBER: US/09/127.227
CURRENT FILING DATE: 1998-07-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1452
TYPE: PRT
ORGANISM: herpesvirus
US-09-127-227-2

Query Match      98.9%  Score 1262; DB 4; Length 1452;
Best Local Similarity 98.7%  Pred. No. 6.7e+129;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKICTTCKLPVWPPTL 60
DB 1215 MSKGEELFTGVVPIVLELGDGVNGHKFSVSGEGSDATYKLTILKFKICTTCKLPVWPPTL 1274
QY 61 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 120
DB 1275 VTITLYGVQVCSRYPDHMKRHDFPKSAMPEGYVQERTIFFKDDGNKYKTRAEVKEGDTIV 1334
QY 121 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 180
DB 1335 NRIELKGIDFKEDGNLGHKLEYNSHNVIYIMADKQKNGIKVNFKIRHNIEIGVGLAD 1394
QY 181 HYQONTPIGDGPVLLPDNNHYLSLTSALSKDNEKRDHNVLEFVTAAGITGHMDELYK 238

```

```

1  RESULT 13
2  US-09-091-042A-2
3  Sequence 2, Application US/09091:42A
4  Patent No. 6455300
5  GENERAL INFORMATION:
6  APPLICANT: The Government of the United States of America
7  as represented by the Secretary
8  Department of Health and Human Services
9  Washington, D.C.
10  Hwu Ph.D., Han
11  Bauer Ph.D., Gordon L.
12  TITLE OF INVENTION: NETWORKS AND COMPOUND-LINKS FOR MONITORING
13  ADDRESSSEE: Needle & Rosenberq
14  STREET: 127 Peachtree Street, Suite 127
15  CITY: Atlanta
16  STATE: Georgia
17  COUNTRY: USA
18  ZIP: 30304
19  COMPUTER READABLE FORM:
20  MEDIUM TYPE: Floppy disk
21  COMPUTER: IBM PC compatible
22  OPERATING SYSTEM: PC-DOS/MS-DOS
23  SOFTWARE: Patent In Release #110, Vers 4.0
24  CURRENT APPLICATION DATA: US/09/091.042A
25  APPLICATION NUMBER: 09-091.042A
26  FILING DATE: 08-Jun-1998
27  CLASSIFICATION: unknown
28  PRIORITY APPLICATION DATA:
29  APPLICATION NUMBER: 62/008,473
30  FILING DATE: 08 Dec 1995
31  ATTORNEY/AGENT INFORMATION:
32  NAME: Selby, Elizabeth

```

1395 HYQNTPIGDPVLLFFNYLSIOSAISKCPNEEDHMLLEFVTAGATTOMDEKX 1492

RESULTS

1 US-08-337-915A-2
 2 Sequence 2, Application US/08337915A
 3 Patent No. 562548
 4 GENERAL INFORMATION:
 5 APPLICANT: Tsien, Roger Y.
 6 APPLICANT: Heim, Roger
 7 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
 8 NUMBER OF SEQUENCES: 2
 9 CORRESPONDENCE ADDRESS:
 10 ADDRESSES: Robbins, Beilker & Gatsch
 11 STREET: 201 No. 562548th Pioneeria Street, Suite 510
 12 CITY: Los Angeles
 13 STATE: California
 14 COUNTRY: USA
 15 ZIP: 90012
 16 COMPUTER READABLE FORM:
 17 MEDIUM TYPE: floppy disk
 18 COMPUTER: IBM PC compatible
 19 OPERATING SYSTEM: PC-DOS/MS-DOS
 20 SOFTWARE: Patent in Release #1-C, Version #1-C
 21 CURRENT APPLICATION DATA:
 22 APPLICATION NUMBER: US/08337915A

[illegible]

Search completed: July 16, 2003, 16:15:30
Job time : 16 secs

GenCore version 3.1.1.6
Copyright (c) 1997 - 2003 CompuGen Ltd.

GM protein: - protein search, using sw model

Run on: July 16, 2003, 16:06:17 : Search time 15 seconds
(without alignments)
456.844 Million cell updates/sec

Title: US-09-967-301-3
Perfect score: 1276
Sequence: 1 MSKGEELFTGVVPLVLLNS.....VLQVYTAIIISGMKILYK 255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 2.5

Searched: 262574 seqs, 29422422 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 250000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents, AA*
1: /cgm2_6/ptodata/1/aa/5A-0005.rep*
2: /cgm2_6/ptodata/1/aa/5A-0005.rep*
3: /cgm2_6/ptodata/1/aa/6A-0005A.rep*
4: /cgm2_6/ptodata/1/aa/6B-0005B.rep*
5: /cgm2_6/ptodata/1/aa/6C-0005C.rep*
6: /cgm2_6/ptodata/1/aa/backfiles.rep*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1258	93.6	238	1	US-08-753-144-2 Sequence 2, Appl
2	1258	98.6	238	2	US-08-679-865-2 Sequence 2, Appl
3	1258	98.6	238	2	US-08-586-876-2 Sequence 2, Appl
4	1258	98.6	238	2	US-08-752-553-2 Sequence 2, Appl
5	1258	98.6	238	3	US-08-753-144-2 Sequence 2, Appl
6	1258	98.6	238	4	US-09-594-353-2 Sequence 2, Appl
7	1258	98.6	238	4	US-09-172-063-2 Sequence 2, Appl
8	1258	98.6	238	4	US-09-263-926-2 Sequence 2, Appl
9	1258	98.6	238	4	US-09-723-432-2 Sequence 2, Appl
10	1258	98.6	238	4	US-09-418-783-1 Sequence 1, Appl
11	1255	96.4	238	3	US-08-893-327-16 Sequence 16, Appl
12	1255	98.4	247	3	US-09-893-327-16 Sequence 18, Appl
13	1254	96.3	238	1	US-08-337-315A-2 Sequence 2, Appl
14	1254	98.3	238	4	US-09-123-530-1 Sequence 2, Appl
15	1254	98.3	238	5	PCT-0595-14562-2 Sequence 2, Appl
16	1253	98.2	239	3	US-08-646-538-2 Sequence 2, Appl
17	1253	98.2	239	5	US-08-503-232-2 Sequence 2, Appl
18	1251	96.0	1070	4	US-09-091-042A-2 Sequence 2, Appl
19	1251	96.0	1452	4	US-09-127-232-2 Sequence 2, Appl
20	1248	97.8	238	3	US-08-643-734A-63 Sequence 49, Appl
21	1245	97.6	247	3	US-08-893-327-20 Sequence 20, Appl
22	1244	97.5	238	4	US-09-213-343-4 Sequence 4, Appl
23	1242	97.3	238	2	US-08-586-261-2 Sequence 2, Appl
24	1242	97.3	238	2	US-09-166-565-2 Sequence 2, Appl
25	1242	97.3	238	3	US-08-893-327-2 Sequence 2, Appl
26	1242	97.3	239	4	US-09-513-783A-2 Sequence 53, Appl
27	1242	97.3	282	4	US-09-513-783A-11 Sequence 14, Appl

26	1242	97.3	287	4	US-09-513-783A-6	Sequence 5, Appl
27	1242	97.3	291	4	US-09-513-783A-10	Sequence 10, Appl
28	1242	97.3	292	4	US-09-513-783A-16	Sequence 16, Appl
29	1242	97.3	295	4	US-09-513-783A-20	Sequence 20, Appl
30	1242	97.3	296	4	US-09-513-783A-12	Sequence 12, Appl
31	1242	97.3	302	4	US-09-513-783A-18	Sequence 18, Appl
32	1242	97.3	302	4	US-09-513-783A-26	Sequence 26, Appl
33	1242	97.3	316	4	US-09-513-783A-26	Sequence 26, Appl
34	1242	97.3	326	4	US-09-513-783A-28	Sequence 28, Appl
35	1242	97.3	347	4	US-09-513-783A-28	Sequence 28, Appl
36	1242	97.3	350	4	US-09-513-783A-30	Sequence 30, Appl
37	1242	97.3	558	4	US-09-513-783A-180	Sequence 180, Appl
38	1242	97.3	558	4	US-09-172-063-3	Sequence 3, Appl
39	1240	97.2	239	4	US-09-513-783A-46	Sequence 46, Appl
40	1240	97.2	239	4	US-09-316-919-4	Sequence 4, Appl
41	1240	97.2	241	4	US-09-062-102-1	Sequence 1, Appl
42	1240	97.2	281	4	US-09-364-946-1	Sequence 1, Appl
43	1240	97.2	284	4	US-09-513-783A-2	Sequence 2, Appl
44	1240	97.2	294	4	US-09-513-783A-2	Sequence 2, Appl
45	1240	97.2	323	4	US-09-172-063-21	Sequence 21, Appl

ALIGNMENTS

US-08-753-144-2
Sequence 2, Application: US/08753143A
GENERAL INFORMATION:
APPLICANT: Tsiem, Roger Y.
APPLICANT: Heilm, Roger
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
FILE REFERENCE: 07257/232003
CURRENT APPLICATION NUMBER: US/08753,143A
CURRENT FILING DATE: 1996-11-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 239
TYPE: PRT
ORGANISM: Aquorea victoria
US-08-753-144-2

5777079

Query Match	98.6%	Score	1258	DB ID	Length	239	
Best Local Similarity	98.7%	Pred. No.	5, 16-123				
Matches	235	Conservative	60	Mismatches	3	Gaps	0
Q	1	MSKGEELFTGVVPLVLLNSGSGSRKCAATYCKLTKECTTKLKYPPPTL	60				
D	1	MSKGEELFTGVVPLVLLNSGSGSRKCAATYCKLTKECTTKLKYPPPTL	60				
Q	61	VTTLSYGVOCPSRYFCHMKRHFERSAMPESYVQERTFEKDDGNYKTRAEVKEGTLV	120				
D	61	VTTLSYGVOCPSRYFCHMKRHFERSAMPESYVQERTFEKDDGNYKTRAEVKEGTLV	120				
Q	121	NRIELKGDPEKEDGNILGHKLKLEYNKSHNYIMADKQKNGIKVNFKIHINIEDSGVOLAD	180				
D	121	NRIELKGDPEKEDGNILGHKLKLEYNKSHNYIMADKQKNGIKVNFKIHINIEDSGVOLAD	180				
Q	161	HYQONTPIGDGVLLPDSNYLSQSALSKDPNEKRDHMYLLSFVTAAGTISKMELEK	238				
D	161	HYQONTPIGDGVLLPDSNYLSQSALSKDPNEKRDHMYLLSFVTAAGTISKMELEK	238				

US-08-679-865-2
Sequence 2, Application: US/08679865
Patent No. 5912137
GENERAL INFORMATION:
APPLICANT: Tsiem, Roger Y.
APPLICANT: Cubitt, Andrew B.
TITLE OF INVENTION: Assays for Protein Kinases Using
Fluorescent Protein Substrates
TITLE OF INVENTION: Fluorescent Protein Substrates
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:


```

1 ADDRESSEE: Townsend and Townsend and Clark LLP
2 STREET: Two Embarcadero Center, Eighth Floor
3 CITY: San Francisco
4 STATE: California
5 COUNTRY: USA
6 ZIP: 94111-3831
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: Patentin Release #1.0, Version #1.0
12 CURRENT APPLICATION DATA:
13 FILING DATE: 16-JUL-1996
14 CLASSIFICATION: 435
15 ATTORNEY/AGENT INFORMATION:
16 NAME: Storella, John S.
17 REGISTRATION NUMBER: 32,944
18 REFERENCE/DOCKET NUMBER: 023072-069000
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: (415) 576-0209
21 TELEFAX: (415) 576-0300
22 INFORMATION FOR SEQ ID NO: 1:
23 SEQUENCE CHARACTERISTICS:
24 LENGTH: 238 amino acids
25 TYPE: amino acid
26 TOPOLOGY: linear
27 MOLECULE TYPE: protein
28 US-08-679-855-2
29
30 Query Match 96.6%; Score 1258; DB 2; Length 238;
31 Best Local Similarity 96.7%; Pred. No. 5,16-129;
32 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
33
34 QY 1 MSKGEELFTGVVPIVLVDSDVNGHFKFSVSGEGGDATYKLTILKFICTTGKLPVWPWT 60
35 DB 1 MSKGEELFTGVVPIVLVDSDVNGHFKFSVSGEGGDATYKLTILKFICTTGKLPVWPWT 60
36
37 QY 61 VTTLSYGVQCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDGNYKTRAEVKFEGDITLV 120
38 DB 61 VTTLSYGVQCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDGNYKTRAEVKFEGDITLV 120
39
40 QY 121 NRIELKGIDKEDNGILGHKLEYNHNHNYIMADKQNGIKVNEKIRHNLEGGVOLAD 180
41 DB 121 NRIELKGIDKEDNGILGHKLEYNHNHNYIMADKQNGIKVNEKIRHNLEGGVOLAD 180
42
43 QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALS KDPNKRDMHVLGVFTAAGITHGMDELYK 238
44 DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALS KDPNKRDMHVLGVFTAAGITHGMDELYK 238
45
46 RESULT 3
47 US-08-680-876-2
48 Sequence 2, Application US/08680876
49 Patent No. 5925558
50 GENERAL INFORMATION:
51 APPLICANT: Tsien, Roger Y.
52 APPLICANT: Cubitt, Andrew B.
53 TITLE OF INVENTION: Assays for Protein Kinase Activity
54 TITLE OF INVENTION: Fluorescent Protein Substrates
55 NUMBER OF SEQUENCES: 48
56 CORRESPONDENCE ADDRESS:
57 ADDRESSEE: Townsend and Townsend and Clark LLP
58 STREET: Two Embarcadero Center, Eighth Floor
59 CITY: San Francisco
60 STATE: California
61 COUNTRY: USA
62 ZIP: 94111-3834
63 COMPUTER READABLE FORM:
64 MEDIUM TYPE: Floppy disk
65 COMPUTER: IBM PC compatible
66 OPERATING SYSTEM: PC-DOS/MS-DOS
67 SOFTWARE: Patentin Release #1.0, Version #1.0

```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/680,876
3 FILING DATE: 16-JUL-1996
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Storella, John S.
7 REGISTRATION NUMBER: 32,944
8 REFERENCE/DOCKET NUMBER: 023072-069200
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (415) 576-0200
11 TELEFAX: (415) 576-0300
12 INFORMATION FOR SEQ ID NO: 2:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 238 amino acids
15 TYPE: amino acid
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 US-08-680-876-2
19
20 Query Match 96.6%; Score 1258; DB 2; Length 238;
21 Best Local Similarity 96.7%; Pred. No. 5,16-129;
22 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
23
24 QY 1 MSKGEELFTGVVPIVLVDSDVNGHFKFSVSGEGGDATYKLTILKFICTTGKLPVWPWT 60
25 DB 1 MSKGEELFTGVVPIVLVDSDVNGHFKFSVSGEGGDATYKLTILKFICTTGKLPVWPWT 60
26
27 QY 61 VTTLSYGVQCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDGNYKTRAEVKFEGDITLV 120
28 DB 61 VTTLSYGVQCFSRYPDHMKRHDFFKSPMPESYVQERTIFFKDGNYKTRAEVKFEGDITLV 120
29
30 QY 121 NRIELKGIDKEDNGILGHKLEYNHNHNYIMADKQNGIKVNEKIRHNLEGGVOLAD 180
31 DB 121 NRIELKGIDKEDNGILGHKLEYNHNHNYIMADKQNGIKVNEKIRHNLEGGVOLAD 180
32
33 QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALS KDPNKRDMHVLGVFTAAGITHGMDELYK 238
34 DB 181 HYQONTPIGDGPVLLPDNHYLSTOSALS KDPNKRDMHVLGVFTAAGITHGMDELYK 238
35
36 RESULT 4
37 US-08-792-555-2
38 Sequence 2, Application US/08792553
39 Patent No. 5981200
40 GENERAL INFORMATION:
41 APPLICANT: Tsien, Roger Y.
42 APPLICANT: Heim, Roger
43 TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
44 NUMBER OF SEQUENCES: 25
45 CORRESPONDENCE ADDRESS:
46 ADDRESSEE: FISH & RICHARDSON P.C.
47 STREET: 4225 Executive Square, Suite 1400
48 CITY: La Jolla
49 STATE: California
50 COUNTRY: USA
51 ZIP: 92037
52 COMPUTER READABLE FORM:
53 MEDIUM TYPE: Floppy disk
54 COMPUTER: IBM PC compatible
55 OPERATING SYSTEM: PC-DOS/MS-DOS
56 SOFTWARE: Patentin Release #1.0, Version #1.25
57 CURRENT APPLICATION DATA:
58 FILING DATE: 31-JAN-1997
59 CLASSIFICATION: 435
60 ATTORNEY/AGENT INFORMATION:
61 NAME: Lisa A. Hall, Ph.D.
62 REGISTRATION NUMBER: 38,347
63 REFERENCE/DOCKET NUMBER: 07257/041001/00 96-160-2
64 TELECOMMUNICATION INFORMATION:
65 TELEPHONE: 619-678-5070
66 TELEFAX: 619-678-5099
67 INFORMATION FOR SEQ ID NO: 2:

```


APPLICANT: Wachter, Rebekka M.

APPLICANT: Reelington, S. James

TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES

FILE REFERENCE: 07257/032001

CURRENT FILING DATE: 1998-10-13

EARLIER FILING DATE: 1998-06-09

NUMBER OF SEQ ID NOS: 34

SOFTWARE: FASTSEQ for Windows Version 1.0

SEQ ID NO 2

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria

US-09-172-053-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

DB 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

RESULT 8

US-09-263-975-2

Sequence 2, Application US/0426-975

Patent No. 6248550

GENERAL INFORMATION:

APPLICANT: Tsieng, Roger Y.

APPLICANT: Cubitt, Andrew H.

TITLE OF INVENTION: Assays for Protein Kinases Using

TITLE OF INVENTION: Fluorescent Protein Substrates

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Clark LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/263,975

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/479,485

FILING DATE: JANU-1996

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John S

REGISTRATION NUMBER: 32,844

REFERENCE/DOCKET NUMBER: 023072-065080

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 238 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-263-975-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

DB 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

RESULT 9

US-08-727-452-2

Sequence 2, Application US/08727452A

Patent No. 6319669

GENERAL INFORMATION:

APPLICANT: Tsieng, Roger Y.

APPLICANT: Helm, Roger

TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES

FILE REFERENCE: 07257/032001

CURRENT APPLICATION NUMBER: US/08/727.452A

CURRENT FILING DATE: 1996-03-20

EARLIER APPLICATION NUMBER: PCT/US95/14692

EARLIER FILING DATE: 1995-11-23

EARLIER APPLICATION NUMBER: US 07/337,915

EARLIER FILING DATE: 1994-11-10

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 2

LENGTH: 238

TYPE: PRT

ORGANISM: Aequorea victoria

US-08-727-452-2

Query Match 98.6% Score 1258; DB 4; Length 238;

Best Local Similarity 98.7% Pred. No. 5,1e-129;

Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

DB 1 MSKGEELFTGVPIVLELGDGVNGHKFVSSEGEADATYKLTILKFKICTGKLPVFWPTL 60

QY 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

DB 61 VTLSYGVOCFSRYPDHKKRHDFKSAPEGYVOERTIFFKDDGNKYKTRAEVKEFEDTLV 120

QY 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

DB 121 NRLELKGIDFKEDGNILGHKLEYNYNHSHNYIMADKOKNGIKVNEKIRHNIEGSGVQLAD 180

QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPNHNYLSTQSALSCKDPNEKRHRHMLLEFVTAAGITHGMDELYK 238

```

RESULTS
US-09-418-785-1
Sequence 1, Application US/094-5785
Patent No. 6414113
GENERAL INFORMATION:
APPLICANT: Fisher, Hugh
TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
FILE OF INVENTION: the Aquorea victoria Green Fluorescent Protein
FILE REFERENCE: RUC 99-0011
CURRENT APPLICATION NUMBER: US/09-418-785
PRIOR APPLICATION NUMBER: 1999-10-15
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 238
TYPE: PRT
ORGANISM: Aquorea victoria
PUBICATION INFORMATION:
AUTHORS: Prasher, D.C. et al.
TITLE: Primary structure of the Aquorea victoria Green-f
JOURNAL: Gene
VOLUME: 112
PAGES: 229-233
DATE: 1992-01-01
DATABASE ACCESSION NUMBER: Genbank M62553
DATABASE ENTRY DATE: 1993-04-26
US-09-418-785-1
Query Match 96.4% Score 12551 DB 3 Length 238
Best Local Similarity 96.3% Pred. No. 11e-128
Matches 234: Conservative 2; Mismatches 2; Indels 0; Gaps 0
QY 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFICTGKLVNPTL 60
DU 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGDATYGLIKFICTGKLVNPTL 60
QY 51 VTLSYGVCCSRYPDKKHDFPKSAMPEGVQERTIFFKDDGNKYTKAEVKESCTLV 120
DB 61 VTLSYGVCCSRYPDKKHDFPKSAMPEGVQERTIFFKDDGNKYTKAEVKESCTLV 120
QY 121 NRLEKGDIFKEDGNIGHKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
DB 121 NRLEKGDIFKEDGNIGHKLEYNNSHNVYIMADKQKNGIKVNFKISHNIEDAGVOLAD 180
QY 181 HYQNTPIGDPGVLLPDNHYLSTQSALSKDNKRDHNVILGFTVTAAGTHGMDELYK 238
DB 181 HYQNTPIGDPGVLLPDNHYLSTQSALSKDNKRDHNVILGFTVTAAGTHGMDELYK 238
RESULTS
US-08-693-327-13
Sequence 16, Application US/08693327
Patent No. 6020192
GENERAL INFORMATION:
APPLICANT: Zolotukhin, Sergei
APPLICANT: Hauswirth, William W.
APPLICANT: Mazuycka, Nicholas
TITLE OF INVENTION: Mutarized Green Fluorescent Protein
TITLE OF INVENTION: Genes and Methods
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893.327
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/586,201
 FILING DATE: 18-CAN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kitchell, Barbara S.
 REGISTRATION NUMBER: 24,528
 REFERENCE/DOCKET NUMBER: 01LA062AKIT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-0660
 TELEFAX: (713) 789-2679
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 247 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-693-327-19

Query Match 98.3% Score 1254; DB 1; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1.4e-128;
 Matches 234; Conservative 1; Mismatches 2; Indels 3; Gaps 0;
 QY 1 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 D5 10 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 QY 61 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 D5 70 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 QY 121 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 D5 130 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 238
 D5 190 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 247

RESULT 13
 US-08-337-915A-2
 Sequence 2, Application US/08337915A
 Patent No. 5625248
 GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Helm, Roger
 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Robbins, Berliner & Carson
 STREET: 201 No. 562504th Figueroa Street, Suite 500
 CITY: Los Angeles
 STATE: California
 COUNTRY: USA
 ZIP: 90012
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/337,915A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Spitals, John P.
 REGISTRATION NUMBER: 29,215
 REFERENCE/DOCKET NUMBER: 1279-178
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 977-1001
 TELEFAX: (213) 977-1003
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-967-301-3.ra1
 Query Match 98.3% Score 1254; DB 1; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1.4e-128;
 Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 D5 1 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 QY 61 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 D5 61 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 QY 121 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 D5 121 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 238
 D5 181 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 238
 RESULT 14
 US-09-121-539-1
 Sequence 1, Application US/09121539B
 Patent No. 6135416
 GENERAL INFORMATION:
 APPLICANT: Osumi, Takashi
 APPLICANT: Tsukamoto, Toshio
 APPLICANT: Tsukamoto, No. 6194545iyo
 APPLICANT: Yamasaki, Masa-oshi
 TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
 TITLE OF INVENTION: PROTEINS
 FILE REFERENCE: 046124-5005
 CURRENT APPLICATION NUMBER: US/09/121,539B
 CURRENT FILING DATE: 1998-07-24
 PRIOR APPLICATION NUMBER: JP 026418/1998
 PRIOR FILING DATE: 1998-01-23
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 1
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea victoria
 FEATURE:
 OTHER INFORMATION: Green fluorescent protein
 US-09-121-539-2

Query Match 98.3% Score 1254; DB 4; Length 238;
 Best Local Similarity 98.3%; Pred. No. 1.4e-128;
 Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 D5 1 MSKGEELFTGVWPLVLELDGVNKHKFSVSGEGDAIYKLTLPKPICTISKLPPWPTL 60
 QY 61 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 D5 61 VTTLTSYGVCFSEYPOHMKRHDFFKSAPEGVYQERTIFFKDGNYKRAEVKFEEDLY 120
 QY 121 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 D5 121 NRTELKGIQFKEDGNILGHKLEYNNSHNVYIMADKQKNSIKVNFKIRHNIEDGSGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 238
 D5 181 HYQONTPIGDGPVLLPNHVLSTQSALSKEPNKRDHVMLEFVTAAGTTHKMEFLYK 238

RESULT 15
PCT-US95-14692-2
Sequence 2: Application: PC/TOS9514692
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carsen
STREET: 201 North Figueroa Street, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent's Release #1.0, Version #1.15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14692
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Spitznagel, John P.
REGISTRATION NUMBER: 29,215
REFERENCE/DOCKET NUMBER: 1279-17#
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 977-1001
TELEFAX: (213) 977-1001
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amio acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-14692-2

Query Match 98.3% Score 1254; DR 51; Decals 212
Best Local Similarity 98.3% Prod. No. 1.40-124
Matches 234; Conservative 1; Misses 10; Gaps 0
CY 1 MSKGEELPTGVVPIVDELQDVNGHREFSVSGGESEFATVYKLTILAPDTTEKLPVWETI 60
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
FR 1 MSKGEELPTGVVPIVDELQDVNGHREFSVSGGESEFATVYKLTILAPDTTEKLPVWETI 60
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 61 VTTLVSVOGCPSSYKGMKSHLFFKSAHTEVYVCEFTTEKKGKNTKTAENKSEFEDLY 120
III IIII IIII IIII IIII IIII IIII IIII IIII
DA 61 VTTFESVOGCPSSYFDMKSHLFFKSAHTEVYVCEFTTEKKGKNTKTAENKSEFEDLY 120
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 121 NRLELKGTECFKEDNGLGKLEKYNYSNHVYIMAKDKNIIKYNKIKRINIEIGAVVLA 180
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
LD 121 NRLELKGTECFKEDNGLGKLEKYNYSNHVYIMAKDKNIIKYNKIKRINIEIGAVVLA 180
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
CY 181 HYQNTPIGPGVLLDNNHVLSTQSALSKEPKNEKRMVLAHVIAAGTIHMDCELYK 238
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII
LD 181 HYQNTPIGPGVLLDNNHVLSTQSALSKEPKNEKRMVLAHVIAAGTIHMDCELYK 238
IIIIII IIII IIII IIII IIII IIII IIII IIII IIII

Search completed: July 16, 2003, 16:15:29
Job time : 16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer Inc.

ORF protein - protein search, using sw model

Run on: July 16, 2003, 16:08:17 : Search time 15 Seconds
(without alignments)
466,844 Million cell updates/sec

Title: us-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELFGVPIHVELDGVNHNKESVSQDEGATYKGLTKFKTCTSKLPVWKPVL 238

Scoring table:

Gapop 16.0, Gapext 2.5

Searched: 242574 seqs, 2342922 residues

Total number of hits satisfying chosen parameters: 242574

Minimum DS seq length: 0

Maximum DH seq length: 200000000

Post-processing: Minimum Match 6%

Display first 45 summaries

Database: Issued Patents,AA*

- 1: /cgn2_6/ptodata/1/aa/5A/098.pep*
- 2: /cgn2_6/ptodata/1/aa/5B/098.pep*
- 3: /cgn2_6/ptodata/1/aa/5C/098.pep*
- 4: /cgn2_6/ptodata/1/aa/5D/098.pep*
- 5: /cgn2_6/ptodata/1/aa/5E/098.pep*
- 6: /cgn2_6/ptodata/1/aa/5F/098.pep*
- 7: /cgn2_6/ptodata/1/aa/5G/098.pep*
- 8: /cgn2_6/ptodata/1/aa/5H/098.pep*
- 9: /cgn2_6/ptodata/1/aa/5I/098.pep*
- 10: /cgn2_6/ptodata/1/aa/5J/098.pep*
- 11: /cgn2_6/ptodata/1/aa/5K/098.pep*
- 12: /cgn2_6/ptodata/1/aa/5L/098.pep*
- 13: /cgn2_6/ptodata/1/aa/5M/098.pep*
- 14: /cgn2_6/ptodata/1/aa/5N/098.pep*
- 15: /cgn2_6/ptodata/1/aa/5O/098.pep*
- 16: /cgn2_6/ptodata/1/aa/5P/098.pep*
- 17: /cgn2_6/ptodata/1/aa/5Q/098.pep*
- 18: /cgn2_6/ptodata/1/aa/5R/098.pep*
- 19: /cgn2_6/ptodata/1/aa/5S/098.pep*
- 20: /cgn2_6/ptodata/1/aa/5T/098.pep*
- 21: /cgn2_6/ptodata/1/aa/5U/098.pep*
- 22: /cgn2_6/ptodata/1/aa/5V/098.pep*
- 23: /cgn2_6/ptodata/1/aa/5W/098.pep*
- 24: /cgn2_6/ptodata/1/aa/5X/098.pep*
- 25: /cgn2_6/ptodata/1/aa/5Y/098.pep*
- 26: /cgn2_6/ptodata/1/aa/5Z/098.pep*
- 27: /cgn2_6/ptodata/1/aa/5[098.pep*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Summary

Result No.	Score	Match	Length	DS	ID	Description
1	1275	100.0	238	1	US-08-679-865-2	Sequence 2, Appl
2	1275	100.0	238	2	US-08-679-865-2	Sequence 2, Appl
3	1275	100.0	238	2	US-08-679-865-2	Sequence 2, Appl
4	1275	100.0	238	2	US-08-679-865-2	Sequence 2, Appl
5	1275	100.0	238	3	US-08-679-865-2	Sequence 2, Appl
6	1275	100.0	238	3	US-08-679-865-2	Sequence 2, Appl
7	1275	100.0	238	4	US-08-679-865-2	Sequence 2, Appl
8	1275	100.0	238	4	US-08-679-865-2	Sequence 2, Appl
9	1275	100.0	238	4	US-08-679-865-2	Sequence 2, Appl
10	1275	100.0	238	4	US-08-679-865-2	Sequence 2, Appl
11	1271	99.7	238	1	US-08-679-865-2	Sequence 2, Appl
12	1271	99.7	238	4	US-08-679-865-2	Sequence 2, Appl
13	1271	99.7	238	5	US-08-679-865-2	Sequence 2, Appl
14	1270	99.6	238	3	US-08-679-865-2	Sequence 2, Appl
15	1270	99.6	238	4	US-08-679-865-2	Sequence 2, Appl
16	1268	99.5	1375	4	US-09-091-042A-2	Sequence 2, Appl
17	1268	99.5	1452	4	US-09-127-227-2	Sequence 2, Appl
18	1262	99.0	238	3	US-08-893-327-16	Sequence 16, Appl
19	1262	99.0	247	3	US-08-893-327-18	Sequence 18, Appl
20	1262	99.0	247	3	US-08-893-327-20	Sequence 20, Appl
21	1261	98.9	238	4	US-08-213-343-4	Sequence 4, Appl
22	1259	98.7	238	2	US-08-588-211-2	Sequence 2, Appl
23	1259	98.7	238	2	US-08-169-659-2	Sequence 2, Appl
24	1259	98.7	238	3	US-08-893-327-2	Sequence 2, Appl
25	1255	98.4	238	3	US-08-643-704A-49	Sequence 49, Appl
26	1253	98.3	239	4	US-09-121-519-14	Sequence 14, Appl
27	1250	98.0	238	1	US-08-452-295-1	Sequence 1, Appl

ALIGNMENTS

5277079

RESULT 1

US-08-679-865-2

Sequence 2, Application US/087511446

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS

FILE REFERENCE: 07257/32093

CURRENT APPLICATION NUMBER: US/08/753,144A

CURRENT FILING DATE: 1996-11-20

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 2

LENGTH: 238

TYPE: PRT

ORGANISM: Neisseria meningitidis

US-08-753-144-2

Query Match: 100.0% Score: 1275; DB: 1; Length: 238;

Host Local Similarity: 100.0%; Pred. No.: 1e-129;

Mismatches: 238; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

1 MSKGEELFGVPIHVELDGVNHNKESVSQDEGATYKGLTKFKTCTSKLPVWKPVL 60

2 MSKGEELFGVPIHVELDGVNHNKESVSQDEGATYKGLTKFKTCTSKLPVWKPVL 60

3 VTFSSYGVQCFRYPDHNKRDHDFKSAMDEGVYQVPTTFKDSGNKTKRAEVKFGDTLV 120

4 VTFSSYGVQCFRYPDHNKRDHDFKSAMDEGVYQVPTTFKDSGNKTKRAEVKFGDTLV 120

5 NRKELKGLDFKEDNLCCHKLEYNNSNNYIMADKOKNKKYKFNKIRHNIEGNSVLAD 180

6 NRKELKGLDFKEDNLCCHKLEYNNSNNYIMADKOKNKKYKFNKIRHNIEGNSVLAD 180

7 HYQNTPTGDPVLLPDNHYSTQSAISKPNKRKHIMVLEFVTAAGITHRMCELYK 238

8 HYQNTPTGDPVLLPDNHYSTQSAISKPNKRKHIMVLEFVTAAGITHRMCELYK 238

RESULT 2

US-08-679-865-2

Sequence 2, Application US/08673865

Patent No. 5912137

GENERAL INFORMATION:

APPLICANT: Tsien, Roger Y.

TITLE OF INVENTION: Assays for Protein Kinases Using

Fluorescent Protein Substrates

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3634
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/579,445
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Storella, John S.
 REGISTRATION NUMBER: 32,944
 REFERENCE/DOCKET NUMBER: 024072-069200
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-679-865-2

Query Match 100.0% Score 1275; Db 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1e-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGREGDATYVKLTLPICITGKLVKPTL 60
 DB 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGREGDATYVKLTLPICITGKLVKPTL 60
 QY 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNGKTRAEVKEEDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNGKTRAEVKEEDTLV 120
 QY 121 NRLEKGIKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNFKIRHNIESVGLAD 180
 DB 121 NRLEKGIKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNFKIRHNIESVGLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSQKPNKRDHMLLEFVTAAGITHGMKIELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQKPNKRDHMLLEFVTAAGITHGMKIELYK 238

RESULT 3
 US-08-580-876-2
 Sequence 2, Application US/58684876
 Patent No. 5925558

GENERAL INFORMATION:
 APPLICANT: Ision, Roger Y.
 APPLICANT: Cubitt, Andrew B.
 TITLE OF INVENTION: Assays for Protein Kinases Using
 TITLE OF INVENTION: Fluorescent Protein Substrates
 NUMBER OF SEQUENCES: 49
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3634
 COMPUTER READABLE FORM:
 MEDIUM TYPE: disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/580,876
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Storella, John S.
 REGISTRATION NUMBER: 32,944
 REFERENCE/DOCKET NUMBER: 024072-069200
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-679-865-2

Query Match 100.0% Score 1275; Db 2; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1e-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGREGDATYVKLTLPICITGKLVKPTL 60
 DB 1 MSKGEELFTGVVFLVELDGVNKHKFSVSGREGDATYVKLTLPICITGKLVKPTL 60
 QY 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNGKTRAEVKEEDTLV 120
 DB 61 VTTFSGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDGNGKTRAEVKEEDTLV 120
 QY 121 NRLEKGIKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNFKIRHNIESVGLAD 180
 DB 121 NRLEKGIKEDGNILGKLEYNYNHNVYIMADKOKNIGKVNFKIRHNIESVGLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSQKPNKRDHMLLEFVTAAGITHGMKIELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSQKPNKRDHMLLEFVTAAGITHGMKIELYK 238

RESULT 4
 US-08-792-553-2
 Sequence 2, Application US/08792553
 Patent No. 5981200
 GENERAL INFORMATION:
 APPLICANT: Ision, Roger Y.
 APPLICANT: Helm, Roger
 TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & RICHARDSON P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: California
 COUNTRY: USA
 ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/792,553
 FILING DATE: 31-JAN-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Lisa A. Hallie, Ph.D.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 07257/041001/UC 96-150-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-678-5070
 TELEFAX: 619-678-5099
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-792-553-2

Query Match 100.0% Score 1275; DB 2; Length 238;
 Best Local Similarity 100.0% Pred. No. 1e-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60

CY 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120
 DB 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120

CY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180

CY 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238

RESULT 5

US-09-753-144-2
 Sequence 2, Application US/09753144
 Patent No. 606476
 GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Heam, Roger
 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Elson A. Richardson P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: CA
 COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09753,144
 FILING DATE: 20-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/727,452
 FILING DATE: 10-OCT-1996
 APPLICATION NUMBER: US95/14692
 FILING DATE: 13-NOV-1995
 APPLICATION NUMBER: 08/337,915
 FILING DATE: 10-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Hall, Lisa A.
 REGISTRATION NUMBER: 38,147
 REFERENCE/POCKET NUMBER: 07257/032002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/478-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal

US-09-753-144-2

Query Match 100.0% Score 1275; DB 3; Length 238;
 Best Local Similarity 100.0% Pred. No. 1e-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60

CY 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120
 DB 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120

CY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180

CY 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238

RESULT 6

US-09-964-359-2
 Sequence 2, Application US/09094359
 Patent No. 614032
 GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Miyawaki, Atsushi
 APPLICANT: Clodis, Juan
 TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 FILE REFERENCE: 07257/067003
 CURRENT APPLICATION NUMBER: US/09/094,359
 CURRENT FILING DATE: 1998-06-09
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 2
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea victoria
 US-09-964,359-2

Query Match 100.0% Score 1275; DB 4; Length 238;
 Best Local Similarity 100.0% Pred. No. 1e-129;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60
 DB 1 MSKGEELFTGVVPLVDELGVNGHKFSVSGEGGATYKLTLCCTGKLPVWFTL 60

CY 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120
 DB 61 VTTFSYGVQCFSRYPDMKRRHDFKFSAMPFGYVQERTIFPKDGNKYKTRAEVKESTLV 120

CY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEGSSVOLAD 180

CY 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238
 DB 181 HYQONTPIGDGVPVLLDPNHVLTQSALSADPNKRRHMHVLEFVTAAGTTHGMCELYK 238

RESULT 7

US-09-172-063-2
 Sequence 2, Application US/09172063
 Patent No. 6150176
 GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Miyawaki, Atsushi
 APPLICANT: Clodis, Juan

```

7 APPLICANT: Wachter, Rebecca M.
7 APPLICANT: Reister, S. James
7 TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
7 TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
7 FILE REFERENCE: 07257/071001
7 CURRENT APPLICATION NUMBER: US/09/272,163
7 CURRENT FILING DATE: 1998-10-13
7 EARLIER APPLICATION NUMBER: 07/041,459
7 EARLIER FILING DATE: 1998-06-03
7 NUMBER OF SEQ ID NOS: 38
7 SOFTWARE: Fast-Seq for Windows Version 4.0
7 SEQ ID NO 2
7 LENGTH: 238
7 TYPE: PRT
7 ORGANISM: Aequorea victoria
7 US-09-172-063-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
DB 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
QY 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

RESULT 5
US-08-727-452-2
7 Sequence 2, Application US/08/27452A
7 Patent No. 639669
7 GENERAL INFORMATION:
7 APPLICANT: Tsielen, Roger Y.
7 APPLICANT: Helm, Roger
7 TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR DETECTION OF ANALYTES
7 FILE REFERENCE: 07257/032031
7 CURRENT APPLICATION NUMBER: US/08/727,452A
7 EARLIER FILING DATE: 1996-03-20
7 EARLIER APPLICATION NUMBER: PCT/US95/14692
7 EARLIER FILING DATE: 1995-11-13
7 EARLIER APPLICATION NUMBER: US 07/337,915
7 EARLIER FILING DATE: 1994-11-10
7 NUMBER OF SEQ ID NOS: 5
7 SOFTWARE: FastSeq for Windows Version 3.0
7 SEQ ID NO 2
7 LENGTH: 239
7 TYPE: PRT
7 ORGANISM: Aequorea victoria
7 US-08-727-452-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
DB 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
QY 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

APPLICANT: Wachter, Rebecca M.
APPLICANT: Reister, S. James
TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
FILE REFERENCE: 07257/071001
CURRENT APPLICATION NUMBER: US/09/272,163
CURRENT FILING DATE: 1998-10-13
EARLIER APPLICATION NUMBER: 07/041,459
EARLIER FILING DATE: 1998-06-03
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-172-063-2

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
DB 1 MSKGEELFTGVWPIVLVDGVNGHKFVSVSSEGGGATYGKTLKFLKLPVWPTL 60
QY 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
DB 61 VTTFSGVQCFSPYPOHMKRHDFKFSAMPEGYVQERTIFFKDGNGYKTRAEVKEGDTLV 120
QY 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
DB 121 NRLEKGIQDKFKEGNLIGHKLEYNYSNHNYYIMADKQKNGIKYVFKIRKINIEGSSVQLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDELYK 238

RESULT 6
US-09-263-975-2
7 Sequence 2, Application US/09/263975
7 Patent No. 6248550
7 GENERAL INFORMATION:
7 APPLICANT: Tsielen, Roger Y.
7 APPLICANT: Cubitt, Andrew B.
7 TITLE OF INVENTION: Assays for Protein Kinases Using
7 TITLE OF INVENTION: Fluorescent Protein Substrates
7 NUMBER OF SEQUENCES: 48
7 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Townsend and Townsend and Crew LLP
7 STREET: Two Embarcadero Center, Eighth Floor
7 CITY: San Francisco
7 STATE: California
7 COUNTRY: USA
7 ZIP: 94111-3834
7 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
7 SOFTWARE: Patent In Release 4.0.0, Version 4.0.0
7 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/09/263,975
7 FILING DATE:
7 CLASSIFICATION:
7 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 08/679,845
7 FILING DATE: 16-JUL-1996
7 ATTORNEY/AGENT INFORMATION:
7 NAME: Sterella, John S.
7 REGISTRATION NUMBER: 32,444
7 REFERENCE/DOCKET NUMBER: 025072-064500
7 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (415) 576-0260
```

```

RESULT 30
US-09-418-785-1
? Sequence 1, Application US/9418745
? Patent No. 6414119
? GENERAL INFORMATION:
? APPLICANT: Fisher, Ruth
? TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
? TITLE OF INVENTION: The Aequorea victoria Green Fluorescent Protein
? FILE REFERENCE: Pats 94-0011
? CURRENT APPLICATION NUMBER: US/99/418,785
? CURRENT FILING DATE: 1999-10-15
? PRIOR APPLICATION NUMBER: 60/104,563
? PRIOR FILING DATE: 1998-10-16
? NUMBER OF SEQ ID NOS: 3
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 238
? TYPE: PRT
? ORGANISM: Aequorea victoria
? PUBLICATION INFORMATION:
? AUTHORS: Prescher, D.C. et al.
? TITLE: Primary structure of the Aequorea victoria green-1
? JOURNAL: Gene
? VOLUME: 211
? PAGES: 229-233
? DATE: 1992-01-01
? DATABASE ACCESSION NUMBER: Genbank M6265
? DATABASE ENTRY DATE: 1993-04-26
? US-09-418-785-1

Query Match 100.0% Score 1275; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 2.7e-129;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEEFTGVVPIVLDGVNGHFKFSVSGEGDGYGKLTAKFCITGKLPVWPTL 60
DB 1 MSKGEEFTGVVPIVLDGVNGHFKFSVSGEGDGYGKLTAKFCITGKLPVWPTL 60

QY 61 VTFPSYGVCCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRASVKFEDTLV 120
DB 61 VTFPSYGVCCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRASVKFEDTLV 120

QY 121 NRLEKGDIDFKEDGNLGHKLEYNVSHNYIMADKQKNGIKVAFKIRENLEHSGVQLAD 180
DB 121 NRLEKGDIDFKEDGNLGHKLEYNVSHNYIMADKQKNGIKVAFKIRENLEHSGVQLAD 180

QY 181 HYQNTPIGDGVPVLPDNYHLSQSALSKDPNEKRDRHVLFFVTAAGITHGMPELYK 238
DB 181 HYQNTPIGDGVPVLPDNYHLSQSALSKDPNEKRDRHVLFFVTAAGITHGMPELYK 238

RESULT 31
US-09-437-915A-2
? Sequence 2, Application US/9447915A
? Patent No. 5625348
? GENERAL INFORMATION:
? APPLICANT: Isien, Roger Y.
? APPLICANT: Heim, Roger
? TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Robbins, Berliner & Carson
? STREET: 201 No. 5625048th Figueroa Street, Suite 500
? CITY: Los Angeles
? STATE: California
? COUNTRY: USA
? ZIP: 90012
? COMPUTER READABLE FORM:
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version: #1.25

Query Match 99.7% Score 1271; DB 4; Length 238;
Best Local Similarity 99.6%; Pred. No. 2.7e-129;
Matches 237; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MSKGEEFTGVVPIVLDGVNGHFKFSVSGEGDGYGKLTAKFCITGKLPVWPTL 60
DB 1 MSKGEEFTGVVPIVLDGVNGHFKFSVSGEGDGYGKLTAKFCITGKLPVWPTL 60

```

```

QY 61 VTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 120
DB 61 VTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 120
QY 121 NRTELKGDIFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 160
DB 121 NRTELKGDIFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 160
QY 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238
DB 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238

RESULT 13
PCT-US95-14692-2
: Sequence 2, Application PC/TUS9514692
: GENERAL INFORMATION:
: APPLICANT: Helms, Roger Y.
: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Robbins, Reilner & Carlson
: STREET: 201 North Flowerda Street, Suite 500
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90012
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/14692
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Spitals, John P.
: REGISTRATION NUMBER: 29,215
: REFERENCE/DOCKET NUMBER: 1279-178
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 977-1001
: TELEFAX: (213) 977-1003
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US95-14692-2

Query Match 99.78; Score 1271; DB 5; Length 238;
Best Local Similarity 99.68; Pred. No. 2, 26-179;
Matches 217; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 60
DB 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 60
QY 61 VTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 120
DB 61 VTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 120
QY 121 NRTELKGDIFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 160
DB 121 NRTELKGDIFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 160
QY 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238
DB 181 HYQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238

```

```

RESULT 14
US-08-646-538-2
: Sequence 2, Application US/08646538
: Patent No. 6027681
: GENERAL INFORMATION:
: APPLICANT: Pavlakis, George A.
: APPLICANT: Galanaris, George A.
: APPLICANT: Stauber, Roland H.
: APPLICANT: Vourdakis, John N.
: TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
: PROTEIN Having Increased Cellular Fluorescence
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 5th Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/02/046,538
: FILING DATE: NO. 6027681 yet assigned
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 015280 249000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 239 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-646-538-2

Query Match 99.68; Score 1270; DB 3; Length 239;
Best Local Similarity 100.0%; Pred. No. 3 5e-129;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 61
DB 2 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 61
QY 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 62
DB 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTIRFKICTTCKLPVWPTLV 62
QY 62 TTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 121
DB 62 TTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 121
QY 63 TTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 122
DB 63 TTTSYGVGCFSPY:DHMKKQDFPKSAMPEGVVCHTFFKDGNGKTKFAVKKVPEFTLV 122
QY 122 RIEKLGIDFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 161
DB 122 RIEKLGIDFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 161
QY 123 RIEKLGIDFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 162
DB 123 RIEKLGIDFKEDGNLQDKLEFYNNSHNVYIMADKQKNGIKVNFKIHNIENKSVLAAD 162
QY 182 YQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238
DB 182 YQONTPIGDPVLLPNNHYLSTQSALSKDPNEKRDHNVLLFEVTAAGITHMDELK 238

RESULT 15
US-08-539-222-2
: Sequence 2, Application US/08503222
: Patent No. 6265548
: GENERAL INFORMATION:
: APPLICANT: Pavlakis, George A.
: APPLICANT: Galanaris, George A.
: APPLICANT: Stauber, Roland H.

```

1 APPLICANT: Vedinakis, John N.
2 TITLE OF INVENTION: Mutant Acquireda Victoria Fluorescent
3 TITLE OF INVENTION: Proteins Having Increased Cellular Fluorescence
4 NUMBER OF SEQUENCES: 37
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: Townsend and Townsend and Crew LLP
7 STREET: Two Embarcadero Center, 8th Floor
8 CITY: San Francisco
9 STATE: California
10 COUNTRY: USA
11 ZIP: 94111-3834
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC Compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.50
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/523,222
19 FILING DATE:
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US/08/646,536
23 FILING DATE:
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Weber, Kenneth A.
26 REGISTRATION NUMBER: 31,679
27 REFERENCE/DOCKET NUMBER: 015230-24,000
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (415) 776-0290
30 TELEFAX: (415) 776-0400
31 INFORMATION FOR SEQ ID NO. 1:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 239 a.a. acid
34 TYPE: amino acid
35 TOPOLOGY: linear
36 MOLECULE TYPE: protein
37 US-09-503-222-2

Query Match 99.6% Score 1270; DB 1; Length 239
Best local similarity 100.0%; Field No. 358 129; Matches 0; 148 00
Matches 237; Conservative 0; Miscellaneous 0; Tolerable 0; 148 00
CY 2 SKDEELTAVVPTAVPTKIVHVKHKSVSQDGHVATVYKLTKEITVTKLELVVWPTLV 41
DB 3 SKDEELTAVVPTAVPTKIVHVKHKSVSQDGHVATVYKLTKEITVTKLELVVWPTLV 42
CY 62 TTSTSVGGPSRYTPDKMKRDEPKSAMPDYVDEETFEETGHNKTRAPVYVQETLNS 121
DB 63 TTSTSVGGPSRYTPDKMKRDEPKSAMPDYVDEETFEETGHNKTRAPVYVQETLNS 122
CY 122 RIELKQIDKEDNITLGHKLEYNKSNHNYVYADKKKQIKRKKRKHKEIKEDSVQLEKH 181
DB 123 RIELKQIDKEDNITLGHKLEYNKSNHNYVYADKKKQIKRKKRKHKEIKEDSVQLEKH 182
CY 182 YOONTFGRGVPVLPDHPNHYLSTQSALSQDKPKRKHMHVLEEVTAAGTTHGMDEYK 236
DB 183 YOONTFGRGVPVLPDHPNHYLSTQSALSQDKPKRKHMHVLEEVTAAGTTHGMDEYK 237

Search completed: July 16, 2003, 16:15:28
Job time : 16 secs

Sequence version 1.0.0
Copyright (c) 1993 - 2003, Copyright Ltd.

AM protein - protein search, using sw model

Run on: July 16, 2003, 16:00:47 : Search time 35 seconds
Total hits 64,748 Million cell updates/sec

Title: US-09-967-301-4
Perfect score: 1276
Sequence: 1 MSKGEELFGVPIVLELDG.....VLRFVYVAGTTCMEPLK 248

Scoring table: BLASTM62
Gapop 10.0, Sextend 0.5

Searched: 905470 seqs, 124250620 residues

Total number of hits satisfying chosen parameters: 905470

Minimum PB seq length: 0
Maximum PB seq length: 2600000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_101002.*			
1:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1992.DAT*		
2:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1991.DAT*		
3:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1990.DAT*		
4:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1989.DAT*		
5:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1988.DAT*		
6:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1987.DAT*		
7:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1986.DAT*		
8:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1985.DAT*		
9:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1984.DAT*		
10:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1983.DAT*		
11:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1982.DAT*		
12:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1981.DAT*		
13:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1980.DAT*		
14:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1979.DAT*		
15:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1978.DAT*		
16:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1977.DAT*		
17:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1976.DAT*		
18:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1975.DAT*		
19:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1974.DAT*		
20:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1973.DAT*		
21:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1972.DAT*		
22:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1971.DAT*		
23:	/SIDS2/accdata/geneseq/geneseq-emb1/AA1970.DAT*		

Prod. No. is the number of results predicted by BLAST. Score is the score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Match	Length	DB ID	Description
1	1270	99.5	238	19	A. victoria green
2	1266	99.2	238	19	Aequorea victoria
3	1266	99.2	238	19	A. victoria green
4	1266	99.2	238	20	Aequorea victoria
5	1266	99.2	238	23	Aequorea victoria
6	1266	99.2	247	20	Aequorea victoria
7	1266	99.2	501	18	Aequorea victoria
8	1262	98.9	238	17	Green fluorescent
9	1262	98.9	238	18	Aequorea victoria
10	1262	98.9	238	18	Aequorea victoria

11	1262	98.9	238	19	AAW76371	A. victoria green
12	1262	98.9	238	19	AAW76106	A. victoria green
13	1262	98.9	238	19	AAW40479	A. victoria green
14	1262	98.9	238	22	AAW73552	Wild-type green fl
15	1262	98.9	238	23	AAW16038	Aequorea victoria
16	1262	98.9	412	23	AAW08630	GFP fusion protein
17	1262	98.9	412	23	AAW08631	GFP fusion protein
18	1262	98.9	412	23	AAW08632	GFP fusion protein
19	1262	98.9	412	23	AAW08633	GFP fusion protein
20	1262	98.9	412	23	AAW08634	GFP fusion protein
21	1262	98.9	554	19	AAW48661	RG fusion protein
22	1262	98.9	554	19	AAW48662	GR fusion protein
23	1262	98.9	655	23	AAW76672	protein related to
24	1262	98.9	926	23	AAW08635	GFP fusion protein
25	1262	98.9	1070	18	AAW17789	Green fluorescent
26	1261	98.9	238	17	AAW05312	Green fluorescent
27	1261	98.9	238	19	AAW76372	A. victoria green
28	1261	98.9	238	19	AAW76110	A. victoria green
29	1261	98.9	238	19	AAW55079	A. victoria green
30	1261	98.9	238	19	AAW55084	A. victoria green
31	1261	98.9	238	23	AAW16043	Aequorea victoria
32	1261	98.9	238	23	AAW16046	Aequorea victoria
33	1260	98.7	238	19	AAW55080	A. victoria green
34	1260	98.7	238	23	AAW16044	Aequorea victoria
35	1260	98.7	238	23	AAW16045	Aequorea victoria
36	1260	98.7	238	23	AAW16047	Aequorea victoria
37	1260	98.7	501	18	AAW31877	GFP variants S65C
38	1260	98.7	501	18	AAW31878	GFP variants S65C
39	1259	98.7	514	18	AAW31876	GFP variants S65C
40	1259	98.7	238	19	AAW76109	A. victoria green
41	1259	98.7	238	19	AAW76111	A. victoria green
42	1259	98.7	238	19	AAW76113	A. victoria green
43	1259	98.7	238	19	AAW61950	Green fluorescent
44	1259	98.7	238	21	AAW33876	Synthetic green fl
45	1259	98.7	238	23	AAW16041	Aequorea victoria

ALIGNMENTS

RESUME 1	
AAW76106	AAW76106 standard; protein: 238 AA.
XX	AAW76106
XX	AAW76106
XX	18 NOV-1998 (first entry)
XX	A. victoria green fluorescent protein mutant F64L/S65T.
XX	Green fluorescent protein: GFP; mutant: jellyfish; excitation: chlorophore.
XX	Aequorea victoria.
XX	Synthetic.
XX	Key
XX	Misc-difference 64
XX	/label= F64L
XX	/note= "Wild-type Phe is replaced with Leu"
XX	Misc-difference 65
XX	/label= S65T
XX	/note= "Wild-type Ser is replaced by Thr"
XX	US580487-A.
XX	06-SEP-1998.
XX	31-JAN-1997.
XX	97US-0791342.
XX	01-FEB-1996: 96US-0010960.
XX	31-JAN-1997: 97US-0791312.

(STED) UNIV LELAND STANFORD JUNIOR.

Corrack RP, Falkow S. Valdivia RB;

WPI: 1998-505643/43.

DNA encoding mutant green fluorescent protein with greater fluorescence intensity than wild-type protein; useful for studying gene expression and protein localization.

Disclosure; Page 7: 15pp; English.

This sequence represents a mutant green fluorescent protein (GFP) from Aequorea victoria in which a leu residue at position 64 in the wild type protein is replaced by a leu and a Ser at position 65 is replaced by a Thr. These mutations occur in the chromophore region. GFP mutants are used in a method to discover GFP's mutated in the chromophore region that fluoresce more brightly than wild-type GFP upon excitation at 488 nm. These mutants can be used in a method for analyzing a cell containing the GFP. The GFP can also be fused to a protein and used to identify the intracellular localization of a protein of interest. A regulatory element could be operatively connected to a coding portion encoding a mutant GFP and exposed to an environmental stimulus. The fluorescence signal from the cell from the cell then measures the effect of the stimulus on the regulatory element.

NOTE: This sequence does not appear in the specification but has been constructed from the wild-type GFP protein represented in AAW65076.

Sequence 238 AA:

Query Match 99.54; Score 1226; ID 18; Length 238;

Best Local Similarity 99.64; Pred. No. 94125;

Katches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKFLIKETKTLVNPPTL 60
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKFLIKETKTLVNPPTL 60
61 VTLLTYGQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDLY 120
61 VTLLTYGQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDLY 120
121 NRLEKIGDKEDGKILGHKLEYNYNHNYIMADKQNGIKYKNEKTEHHNIDGVQLAD 180
121 NRLEKIGDKEDGKILGHKLEYNYNHNYIMADKQNGIKYKNEKTEHHNIDGVQLAD 180
181 HYQONTPIGQGPVLLPDNHYLSTQSALSKDPNKRDRHVVLEFVTAAGTGHGDELK 238
181 HYQONTPIGQGPVLLPDNHYLSTQSALSKDPNKRDRHVVLEFVTAAGTGHGDELK 238

RESULT 2

AAW65076

ID AAW22101 standard; Protein: 238 AA.

AC AAW22101;

XX 10-MAR-1998 (first entry)

DE Aequorea victoria protein GFP10 mutant (65L-S65L).

XX Green fluorescent protein; GFP; jellyfish; GFP; gene; reporter;

XX humanise; expression; mutant.

XX Aequorea victoria.

XX Key Location/Qualifiers

XX Misc-difference 64 /Label: F64L

XX /Note: "Wild-type P64 is replaced by Ser"

XX Misc-difference 65 /Label: S64T

XX /Note: "Wild-type Ser is replaced by Thr"

XX W09726333-A1.

XX 24-JUL-1997.

XX 17-JAN-1997; 97NO US06755.

XX 19-JAN-1996; 96US-0588201.

XX (CYFI) UNIV FLORIDA RES FOUND INC.

XX Hauswirth WW, Kozyczka N, Zolotukhin S;

XX WPI: 1997-185337/35.

Humanised green fluorescent protein gene - optimised to provide high level expression in mammalian cells, used e.g. to label or identify cells, to locate proteins, etc.

XX Claim 115; Page 1: 158pp; English.

This sequence represents a mutant form of the green fluorescent protein GFP10 i.e. F64L-S65T. Green fluorescent proteins (GFP's) could be used as reporter molecules as they absorb blue light and emit green light without requiring any cofactors, substrates, or additional gene products allowing GFP detection in living cells providing meaningful gene expression is achieved. By providing humanised GFP, adapted for expression in mammalian and human cells, problems associated with wild-type jellyfish GFP e.g. variable and low expression levels should be overcome.

Note: The present sequence does not appear in the specification. It has been made by modifying the wild-type GFP10 amino acid sequence found in AAW31295.

Sequence 238 AA:

Query Match 99.24; Score 1266; ID 18; Length 238;

Best Local Similarity 99.24; Pred. No. 134;

Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKFLIKETKTLVNPPTL 60
1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYGLKFLIKETKTLVNPPTL 60
61 VTLLTYGQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDLY 120
61 VTLLTYGQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDLY 120
121 NRLEKIGDKEDGKILGHKLEYNYNHNYIMADKQNGIKYKNEKTEHHNIDGVQLAD 180
121 NRLEKIGDKEDGKILGHKLEYNYNHNYIMADKQNGIKYKNEKTEHHNIDGVQLAD 180
181 HYQONTPIGQGPVLLPDNHYLSTQSALSKDPNKRDRHVVLEFVTAAGTGHGDELK 238
181 HYQONTPIGQGPVLLPDNHYLSTQSALSKDPNKRDRHVVLEFVTAAGTGHGDELK 238

RESULT 3

AAW65076

TC AAW65076 standard; Protein: 238 AA.

XX AAW65076;

XX 15-SEP-1998 (first entry)

XX A. victoria green fluorescent protein variant S65T.

XX Green fluorescent protein; protein kinase; phosphate donor;

XX phosphorylation; drug screening; receptor-ligand binding; signalling;

XX protein-protein interaction; kinase activation.

XX Aequorea victoria.

XX Synthetic.

181 HQQNTPIGDGFVLLPUNHYLSTQSAISKDPNEKHQKVVLESTAAQIIRHMEIYK 238

RESULT 5
AAE16042
ID AAE16042 standard; Protein: 238 AA.
AC
XX AAE16042:
DT 26-MAR-2002 (first entry)
DE Aequorea victoria GFP mutant protein (S67)
KW Green fluorescent protein; cell lineage tracer; protein localization
KW GFP; fusion tag; gene expression marker; fluorescent energy acceptor;
KW immunosay; hybridisation assay; fluorescent energy donor; biosensor;
KW FRET; fluorescence resonance energy transfer; mutant; protein.
XX
XX Aequorea victoria.
CS Synthetic.
XX
XX Key location/qualifiers
XX Misc-difference 45
XX Z/dna- "wild type Ser substituted with Thr"
XX
XX W0202190147-A2
XX 29-NOV-2001.
XX
XX 17-MAY-2001; 2004WO 0516149
XX
XX 19-MAY-2000; 2000US-0575847.
XX (CYBR-) UNIV OREGON STATE.
XX
XX Wachter E. Remington SJ.
XX
XX WPI: 2002-083084/11.
XX
XX Now long wavelength engineered fluorescent proteins, useful as markers
XX for gene expression, tracers of cell lineage or as fusion tags to
XX monitor protein localization, or in detection assays, and immunoassays
XX or hybridization assays
XX
XX Claim 2; Page 1; 16pp; English.

The patent discloses functional engineered fluorescent protein and its
corresponding polynucleotide. The amino acid sequence of the engineered
protein is identical to Aequorea green fluorescent protein (GFP). The
engineered fluorescent proteins of the invention have varied fluorescent
properties and have the ability to respond to ion concentrations via a
change in fluorescent characteristics. They are useful as markers for
gene expression, tracers of cell lineage or as fusion tags to monitor
protein localization within living cells. The engineered fluorescent
proteins are particularly useful for coupling engineered fluorescent
proteins to antibodies, nucleic acids or other receptors for use in
detection assays, e.g. immunoassays or hybridisation assays. They are
useful for tracking the movement of proteins in cells or in systems
for detecting induction of transcription and for the simultaneous
measurement of two or more processes within cells. Proteins of the
invention are also useful as fluorescent energy donors or acceptors
as well as biosensors for detecting amino acids. They are also useful in
fluorescence resonance energy transfer (FRET). The crystal structure
of the GFP is useful for designing mutants having altered fluorescent
characteristics which are particularly useful to identify amino acids
whose substitution alters fluorescent properties of the protein. The
crystal structure of the GFP is also useful for designing mutants having
altered amino acid binding characteristics which are particularly useful for
identifying amino acids whose substitution alters the specificity and
affinity of the binding site to various amino acids, and for identifying amino
acid binding and therefore the concentration of the amino acid. The present
sequence is Aequorea victoria GFP mutant protein (S67).
Note: This sequence is not shown in the specification but is derived

from Aequorea victoria wild type green fluorescent protein shown in
Figure 3 of the specification (AAE16039).

XX
XX
XX Sequence 218 AA:
Query Match 99.2% Score 1266; DB 23; Length 238;
Best Local Similarity 99.2%; Pos. No. 1,30-124;
Matches 236; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFTQVPTLAVELLDGVNCHKEFVSQSGHGSDATYCKLIKFKICTTGKLVWPTL 60
DE 1 MSKGEELFTQVPTLAVELLDGVNCHKEFVSQSGHGSDATYCKLIKFKICTTGKLVWPTL 60
QY 61 VTITTYGVQCFSEYDPRHKKHDFKSAPEGYVCEITTFPKDQSNYKTRAEYAFEGTIV 120
DE 61 VTITTYGVQCFSEYDPRHKKHDFKSAPEGYVCEITTFPKDQSNYKTRAEYAFEGTIV 120
QY 121 NRIELKGIDKEKGGNLDGKHLEYNYSNHYVTMDKQKNGIKYFKIRHNTENGVSULAD 180
DE 121 NRIELKGIDKEKGGNLDGKHLEYNYSNHYVTMDKQKNGIKYFKIRHNTENGVSULAD 180
QY 181 HQQNTPIGDGFVLLPUNHYLSTQSAISKDPNEKHQKVVLESTAAQIIRHMEIYK 238
DE 181 HQQNTPIGDGFVLLPUNHYLSTQSAISKDPNEKHQKVVLESTAAQIIRHMEIYK 238

RESULT 6
AAW96329
ID AAW96329 standard; Protein: 217 AA.
AC
XX AAW96329:
DT 26-JUN-1999 (first entry)
DE Humanised green fluorescent protein
XX
XX Green fluorescent protein; gfp; jellyfish; Aequorea victoria;
XX humanisation; reporter gene; substrate; cofactor; beta galactosidase;
XX firefly luciferase; alkaline phosphatase;
XX chloramphenicol acetyltransferase; CAT; beta glucuronidase; GUS.
XX Synthetic.
XX Aequorea victoria.
XX
XX W030303997-A1.
XX
XX 28-JAN-1999.
XX
XX 16-JUL-1998; 98WO-US14692.
XX
XX 16-JUL-1997; 97US-0833327.
XX (UYFL) UNIV FLORIDA.
XX
XX Hauswirth W, Huzyczka N, Zolotukhin S;
XX WPI: 1999-132241/11.
XX N-PSDB; AAX08455.
XX
XX Humanised green fluorescent protein - used to measure gene
XX expression and identify transformed cells
XX
XX Claim 5; Page 142; 152pp; English.

Humanised green fluorescent protein (gfp) genes can be used to
identify transformed cells, to measure gene expression in vitro and
in vivo, to label specific cells in multicellular organisms (e.g. to
study cell lineage's), to label and locate fusion proteins, and to
study intracellular trafficking. Commonly used reporter genes include
beta-galactosidase, firefly luciferase, alkaline phosphatase;
chloramphenicol acetyltransferase (CAT), and beta glucuronidase
(GUS). However, these have limitations in their use. Frequently,
these reporter genes require the addition of a substrate and the

CC size of certain proteins means that the expression of reporter
 CC fusion proteins can be difficult. The light stimulated GFP
 CC fluorescence is species independent and does not require any
 CC cofactors substrates or additional gene products from Aquorea
 CC victoria as the GFP genes have been homologous, they are
 CC expressed at sufficient levels to be detectable in human cells
 CC unlike previous GFP proteins.

XX XX Sequence 247 AA:
 Query Match 99.24; Score 1256; DR 20; Length 501
 Best Local Similarity 99.24; Prod. No. 3.9e-124;
 Matches 236; Conservative 1; Mismatches 1; Indels 0; Gaps 0

QY 1 MSKGEELFTGVVILVRLDGVNKHKFSVSGEGDGYGKLTGKLTGKLVFWPPTL 60
 DB 10 MSKGEELFTGVVPLVLDGVNKHKFSVSGEGDGYGKLTGKLTGKLVFWPPTL 60
 QY 51 VTTLTYGVQCFSRYPDHMKHDFKKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 DB 7C VTTLTYGVQCFSRYPDHMKHDFKKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120
 QY 121 NRLELKGIDPKKDGNIHLGHLEYNKSHNYVIMADQKNGIKVNEKTPKNEZEGVQLAD 180
 DB 130 NRLELKGIDPKKDGNIHLGHLEYNKSHNYVIMADQKNGIKVNEKTPKNEZEGVQLAD 180
 QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPKNEKRDHMYLLEFVTAAGTTHGMDELYK 248
 DB 190 HYQONTPIGDGPVLLPDNHYLSTQSALSADPKNEKRDHMYLLEFVTAAGTTHGMDELYK 247

RESULT 7
 ID AAW31879 standard; Protein: 501 AA.

XX AC AAW31879;

XX DT 03-FEB-1998 (first entry)

XX D5 GFP variants S65T and W7 tandem fluorescent protein construct.

XX North West Pacific jellyfish; green fluorescent protein, GFP; S65T
 KW fluorescent resonance energy transfer; FRET; enzymatic assay; W7;
 KW enzyme inhibitor; enzyme promoter; PCR primers; protease cleavage sites
 KW tandem fluorescent protein construct; blue fluorescent protein.

XX OS Synthetic.

XX OS Aquorea victoria.

XX Key Generation/Qualifiers

XX Protein 1-248

XX FT Label= S65T_GFP_variant

XX FT Misc-difference 65

XX FT Label= S65T

XX FT /note= "wild type Ser substituted with Thr"

XX FT Peptide 239..264

XX FT Label= linker_moiety

XX FT Cleavage-site 250..251

XX FT /label= trypsin_cleavage_site

XX FT Cleavage-site 253..254

XX FT /label= calpain_cleavage_site

XX FT Cleavage-site 258..259

XX FT /label= trypsin_enterokinase_cleavage_site

XX FT Protein 265..501

XX FT /label= W7_SFP_variant

XX FT Misc-difference 329

XX FT /label= Y66W

XX FT /note= "wild type Tyr substituted with Trp"

XX FT Misc-difference 409

XX FT /label= N146I

XX FT /note= "wild type Asn substituted with Leu"

XX FT Misc-difference 416

XX FT /label= M153T

FT Misc-difference 425
 FT /label= V153A
 FT /note= "wild type Val substituted with Ala"
 FT Misc-difference 475
 FT /label= N212K
 FT /note= "wild type Asn substituted with Lys"

XX W09728241-A1.

XX PN

XX 07-AUG-1997.

XX 31-JAN-1997.

XX 31-JAN-1996.

XX 96US-0594575.

XX (AURO-) AURORE BIOSCIENCES CORP.

XX (REGC) UNIV CALIFORNIA.

XX Subitt A. Helm S. Tsien BY:

XX WPI: 1997-402615/37.

XX Tandem fluorescent protein constructs - have donor and acceptor

XX moieties exhibiting fluorescent linked via cleavable peptide linker.

XX useful in enzymatic assays

XX Claim 3; Page : 88pp; English.

XX This protein sequence is that of a novel tandem fluorescent protein

XX construct, made using Aquorea victoria (North West Pacific jellyfish)

XX green fluorescent protein (GFP) variants S65T and W7. W7 fluoresces

XX at a shorter wavelength than GFP. The construct comprises a donor

XX (e.g. S65T) and an acceptor (e.g. W7) fluorescent protein moiety

XX (donors and acceptors can be green or blue fluorescent proteins), and a

XX linker coupling them. Preferably, the donor is positioned at the

XX N-terminus of the polypeptide relative to the acceptor. The linker moiety

XX is a peptide 5-50 amino acids in length containing a protease cleavage

XX site. In this example, the linker moiety contains many recognition sites

XX for proteases, including trypsin, calpain and enterokinase. The donor and

XX acceptor moieties exhibit fluorescent resonance energy transfer (FRET)

XX when they are cleaved. The constructs are used in enzymatic assays and

XX can be used to isolate new enzymes or enzyme inhibitors or promoters. The

XX specific activity of enzyme (in vivo and in vitro) and compounds altering

XX enzyme activity can be obtained. FRET and hence activity of specific

XX compounds is measured from the acceptor or donor moiety of specific

XX compounds using a ratio between the two.

XX Note: The present sequence does not appear in the specification; it

XX has been made by modifying the native GFP sequence, and adding the

XX linker moiety in the appropriate place.

XX XX Sequence 501 AA:

XX Query Match 99.24; Score 1256; DR 18; Length 501

XX Best Local Similarity 99.24; Prod. No. 3.9e-124;

XX Matches 236; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY 1 MSKGEELFTGVVILVRLDGVNKHKFSVSGEGDGYGKLTGKLTGKLVFWPPTL 60

DB 1 MSKGEELFTGVVPLVLDGVNKHKFSVSGEGDGYGKLTGKLTGKLVFWPPTL 60

QY 61 VTTLTYGVQCFSRYPDHMKHDFKKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120

DB 61 VTTLTYGVQCFSRYPDHMKHDFKKSAMPEGYVOERTIFFKDGNYKTRAEVKEGTLV 120

QY 121 NRLELKGIDPKKDGNIHLGHLEYNKSHNYVIMADQKNGIKVNEKTPKNEZEGVQLAD 180

DB 121 NRLELKGIDPKKDGNIHLGHLEYNKSHNYVIMADQKNGIKVNEKTPKNEZEGVQLAD 180

QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPKNEKRDHMYLLEFVTAAGTTHGMDELYK 238

DB 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPKNEKRDHMYLLEFVTAAGTTHGMDELYK 238

RESULT 8
AAW05304 standard; Protein: 238 AA.
XX
XX
XX 02-APR-1997 (first entry)
XX Green fluorescent protein.
XX
XX Green fluorescent protein: GFP; jellyfish; Aequorea victoria; sea pansy;
KW Renilla reniformis; differential gene expression; protein localization;
KW gene expression tracking; fluorescence.
XX
XX Aequorea victoria.
XX
XX W05623810-A1.
XX 08-AUG-1996.
XX
XX 13-NOV-1995; 95W0-US14434.
XX 10-NOV-1994; 94US-0137315.
XX (REGC) UNIV CALIFORNIA.
XX PA
XX Reia R. Tsien EV;
XX
XX WPI: 1956-37137/57.
XX N-PSDB: AAW35594
XX
XX New modified Aequorea green fluorescent polypeptides - having
PT amino acid changes to provide prods. which exhibit different
PI excitation and emission spectra
XX
XX Disclosure: Page 20-21; 4pp; English.
XX
XX This sequence represents the wild type Aequorea victoria (jellyfish)
CC green fluorescent protein (GFP). The fluorescence of this protein is
CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
CC the one absorbance peak seen in the related GFP from the sea pansy
CC (Renilla reniformis). This protein is modified to provide the GFPs of
CC the invention. The modifications were created by subjecting the cDNA
CC encoding this sequence to site directed mutagenesis using mutagenic PCR
CC primers, or hydroxylamine treatment. The GFPs of the invention are
CC modified to lead to the formation of products with markedly different
CC excitation and emission spectra. Visibly distinct colours, and increased
CC intensities of emission make these products useful in a wide variety of
CC contexts, such as tracking of differential gene expression and protein
CC localisation. The mutations can also be created to modify the encoded
CC GFP so that it only possesses one absorbance peak.
XX
XX Sequence 238 AA:
SQ Query Match 98.9%; Score 1262; DB 17; Length 248;
Best Local Similarity 98.7%; Pred. No. 3.4e-124;
Matches 235; Conservative 1; Mismatches 2; Indels 6; Gaps 0;
QY 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGEDATYKLTIFCTCTKLPVPWPTL 60
DB 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGEDATYKLTIFCTCTKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFDGNYKTPAEYKFPDZLV 120
DB 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFDGNYKTPAEYKFPDZLV 120
QY 121 NRLEKLGIDFKEDGNLGGKLEYNYNNSHNYINAKSKQKNTKVFRTKHNLEGGVGLAD 180
DB 121 NRLEKLGIDFKEDGNLGGKLEYNYNNSHNYINAKSKQKNTKVFRTKHNLEGGVGLAD 180
QY 181 HYQONTPIGDGPVILPONHYLSTGALSCKDPRKRDHNVLLFVTAACITHGMBELYK 238

DB 181 HYQONTPIGDGPVILPONHYLSTGALSCKDPRKRDHNVLLFVTAACITHGMBELYK 238
RESULT 9
AAW22097
XX
XX AAW22097 standard; Protein: 233 AA.
XX
XX AAW22097;
XX 10-MAR-1998 (first entry)
XX Aequorea victoria protein GFP10 mutant S65T.
XX Green fluorescent protein: GFP; jellyfish; gfp10 gene; reporter;
KW humanise; expression; mutant.
XX
XX Aequorea victoria.
XX
XX Key Location/Qualifiers
XX Misc-difference 55 /Label= S65T
XX /Note= "Wild-type Ser is replaced by Thr"
XX
XX W09726330-A1.
XX 24-JUL-1997.
XX
XX 17-JAN-1997; 97W0-US00755.
XX 16-JAN-1996; 96US-0580201.
XX (BYEL) UNIV FLORIDA RES FOUND INC.
XX Hauswirth W, Muzyczka N, Zolotukhin S;
XX WPI: 1997-36537/15.
XX N-PSDB: AAW75698.
XX
XX Humanised green fluorescent protein gene - optimised to provide high
PT level expression in mammalian cells, used e.g. to label or identify
PT cells, to locate proteins, etc.
XX
XX Claim 3; Page 1; 158pp; English.
XX
XX This sequence represents a mutant form of the green fluorescent protein
CC GFP10 in which the Ser at position 65 of the wild-type protein has
CC been replaced by Thr. Green fluorescent proteins (GFP's) could be used
CC as reporter molecules as they absorb blue light and emit green light
CC without requiring any cofactors, substrates, or additional gene products
CC allowing GFP detection in living cells providing meaningful gene
CC expression is achieved. By providing humanised GFP, adapted for
CC expression in mammalian and human cells, problems associated with
CC wild-type jellyfish GFP e.g. variable and low expression levels should be
CC overcome.
CC Note: The present sequence does not appear in the specification. It
CC has been made by modifying the wild-type GFP10 amino acid sequence found
CC in AAW31295.
XX
XX Sequence 235 AA:
SQ Query Match 98.9%; Score 1262; DB 18; Length 238;
Best Local Similarity 98.7%; Pred. No. 3.4e-124;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGEDATYKLTIFCTCTKLPVPWPTL 60
DB 1 MSKGEELFGVGVPIVELDGVNCHKFSVSGEGEDATYKLTIFCTCTKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFDGNYKTPAEYKFPDZLV 120
DB 61 VTTLTYGVQCFSRYPDHMKRPHDFKSAPEGVVQERTIFFKFDGNYKTPAEYKFPDZLV 120

QY 121 NRLEKGLDTFEDQNLGHKLEYNNSHNYIMADKQKNSKIPNKRCHMVLEFVTAAGTTHGMVELYK 238
 DB 121 NRLEKGLDTFEDQNLGHKLEYNNSHNYIMADKQKNSKIPNKRCHMVLEFVTAAGTTHGMVELYK 238
 QY 131 HQONTPIGDSGVILPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120
 DB 131 HQONTPIGDSGVILPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120

RESULT 10

AAW24232
 ID AAW24232 standard; Protein: 238 AA.
 AC AAW24232
 XX

QY 02-FEB-1998 (first entry)

XX Aequorea victoria green fluorescent protein (GFP).
 KW North West Pacific jellyfish; green fluorescent protein; GFP; 9600.
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; tandem fluorescent protein constructs;
 KW blue fluorescent protein; protease cleavage site.

XX Aequorea victoria.

OS W09726261-A1.
 PN

XX W09726261-A1.
 PD

XX 07-AUG-1997.
 PF

XX 31-JAN-1997; 97WO-0501457.
 PR

XX 31-JAN-1998; 98US-0594575.
 PR

XX (AUCG-) Aequorea victoria; GFP.
 PA

XX (REG-) UNIV CALIFORNIA.
 PA

XX Cabibb A, Heim R, Tsien RY.
 PI

XX WP21: 1997-402515/17.
 LR

XX N-PSDB: AAT77541.
 DR

XX Tandem fluorescent protein constructs - have been used to generate
 PT molecules exhibiting fluorescent linked via cleavable peptide linker
 PT useful in enzymatic assays

XX Disclosure: Fig 1: 8pp; English.

XX This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. GFP, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of
 CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-56 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or may be obtained using a ratio between the two.

XX Sequence 238 AA:

Query Match 98.98; Score 1262; DB 19; Length 238;
 Best Local Similarity 98.78; Pred. No. 3,4e-124;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNCHKFSVSGEGDATYCKLTKLTKFCITGKLFVWPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNCHKFSVSGEGDATYCKLTKLTKFCITGKLFVWPTL 60

QY 61 VTTLTYGVQCFSRYPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120
 DB 61 VTTLTYGVQCFSRYPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120
 QY 121 NRLEKGLDTFEDQNLGHKLEYNNSHNYIMADKQKNSKIPNKRCHMVLEFVTAAGTTHGMVELYK 238
 DB 121 NRLEKGLDTFEDQNLGHKLEYNNSHNYIMADKQKNSKIPNKRCHMVLEFVTAAGTTHGMVELYK 238
 QY 181 HQONTPIGDSGVILPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 238
 DB 181 HQONTPIGDSGVILPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 238

RESULT 11

AAW76371
 ID AAW76371 standard; Protein: 238 AA.

XX AAW76371

XX 03-DEC-1998 (first entry)

XX A. victoria green fluorescent protein.

XX Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.

XX Aequorea victoria.

OS JPL0244382-A.
 IN

XX 08-SEP-1998.
 PD

XX 27-FEB-1997; 97JP-0362370.
 PF

XX 27-FEB-1997; 97JP-0362370.
 PR

XX (CHINA) DINAHKKO KENKYUSHO KK.
 PA

XX KFI: 1998-545945/46.
 LR

XX N-PSDB: AAW61751.
 DR

XX Green fluorescent protein - useful for determining promoter and
 PT targeting activity in a cell

XX Disclosure: Page 6-7: 11pp; Japanese.

XX This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from Aequorea victoria. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 6 times
 CC higher than the conventional fluorescent proteins. The nucleic acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested,
 CC comprising steps of introducing the above vector to a host cell and
 CC detecting the fluorescence emitted from the cell. They can additionally
 CC be used for detecting the targeting activity in the cell of an amino
 CC acid sequence to be tested in which the above fluorescent protein is
 CC introduced to a cell and the distribution of the fluorescent protein in
 CC the cell is observed.

XX Sequence 238 AA:

Query Match 98.98; Score 1262; DB 19; Length 238;
 Best Local Similarity 98.78; Pred. No. 3,4e-124;
 Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNCHKFSVSGEGDATYCKLTKLTKFCITGKLFVWPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNCHKFSVSGEGDATYCKLTKLTKFCITGKLFVWPTL 60

QY 61 VTTLTYGVQCFSRYPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120
 DB 61 VTTLTYGVQCFSRYPDKHKKRDEDFKSNPEGYVOERTIPFKDDGNKYKTRAEVKEGDTLV 120

QY 121 NRLEKLGIDFKEDGNILGKLEFYNTNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGKLEFYNTNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 QY 181 HYQONTFGDGPVLLPDNHYLSTQSAISKDNKPKRHHVLEFVTAAGITHGMEELAK 238
 DB 181 HYQONTFGDGPVLLPDNHYLSTQSAISKDNKPKRHHVLEFVTAAGITHGMEELAK 238

RESULT 12

AAW76105
 ID AAW76105 standard: protein: 238 AA.

XX AC AAW76105;

XX DT 18-NOV-1998 (first entry)

XX DE A. victoria green fluorescent protein.

XX KW Green fluorescent protein; GFP; mutants; jellyfish; excitat. ion; chromophore.

XX OS Aequorea victoria.

XX PN CS5804347-A.

XX PD 08-SEP-1998.

XX PF 31-JAN-1997; 97NS-0791322.

XX PR 01-FEB-1996; 96US-0606800.

XX PR 31-JAN-1997; 97US-0791332.

XX PA (STRD) UNIV LELAND STANFORD JUNIOR.

XX PI Cormack RP, Parkow S, Vaidya RH;

XX DR WPI: 1998-505643/43.

XX PT DNA encoding mutant green fluorescent pigment proteins - with greater fluorescence intensity than wild-type proteins, useful for studying gene expression and protein localisation

XX PS Disclosure: Page 1: 15pp; English.

XX CC This sequence represents the green fluorescent protein (GFP) from Aequorea victoria. This sequence is used in a method to discover GFP's mutated in the chromophore region that fluoresce more brightly than wild-type GFP upon excitation at 488 nm. These mutants can be used in a method for analyzing a cell, containing the GFP. The GFP can also be used to a protein and used to identify the intracellular localisation of a protein of interest. A regulatory element could be operatively connected to a coding portion encoding a mutant GFP and exposed to an environmental stimulus. The fluorescence signal from the GFP then measures the effect of the stimulus on the regulatory element.

XX CC Note: This sequence does not appear in the specification but has been retrieved from the EMBL record contained in W55341.

XX SQ Sequence 238 AA;

Query Match 98.9%; Score 1262; DB 19; Length 238;
 Best Local Similarity 98.7%; Pred. No. 3.4e-124;
 Matches 238; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVWTHVELLDGNGHFKFSVSGRGEATVYKTHAEFTCTTKRIPWAFLL 60

DB 1 MSKGEELFTGVWTHVELLDGNGHFKFSVSGRGEATVYKTHAEFTCTTKRIPWAFLL 60

QY 61 VTLTITGVQCFSPYPOHMKRHHDFKSAHEGYVQGTTFPKDKNYKTRAAVKKFEDTLV 120

DB 61 VTFYSYGVQCFSPYPOHMKRHHDFKSAHEGYVQGTTFPKDKNYKTRAAVKKFEDTLV 120

QY 121 NRLEKLGIDFKEDGNILGKLEFYNTNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGKLEFYNTNSHNHYIMADKOKNGKVNFKIRHNIEDGSGVOLAD 180
 QY 181 HYQONTFGDGPVLLPDNHYLSTQSAISKDNKPKRHHVLEFVTAAGITHGMEELAK 238
 DB 181 HYQONTFGDGPVLLPDNHYLSTQSAISKDNKPKRHHVLEFVTAAGITHGMEELAK 238

RESULT 13

AAW40478
 ID AAW40478 standard: protein: 249 AA.

XX AC AAW40478;

XX DT 15-SEP-1998 (first entry)

XX DE A. victoria green fluorescent protein.

XX KW Green fluorescent protein; protein kinase; phosphate donor; phosphorylation; drug screening; receptor-ligand binding; signalling; protein-protein interaction; kinase activation.

XX OS Aequorea victoria.

XX PF Key Location/Qualifiers
 PT Protein 1-238
 PT Note "Partial sequence"

XX WK CS5804351-A;

XX WK 22-JAN-1998.

XX PF 16-JUL-1997; 97WC-0512410.

XX PF 16-JUL-1996; 96US-0608872.

XX PF 16-JUL-1996; 96US-0679850.

XX PF 16-JUL-1996; 96US-0608876.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Cubitt AB, Tsien RY;

XX DR WPI: 1998-110616/10.

XX DR N-PSDB: AAW1125.

XX PT Assays for protein kinase and modulators - using a fluorescent protein substrate which exhibits a different fluorescent property in the phosphorylated and un-phosphorylated state

XX PS Disclosure: Fig 3: 65pp; English.

XX CC This sequence represents a green fluorescent protein from Aequorea victoria which is used in a method to determine if a sample contains protein kinase (PK) activity. The method involves contacting the sample with a phosphate donor (2D) and a fluorescent protein (EP) substrate; a PK, the protein substrate comprising a PK moiety and a phosphorylation site for a PK, where the protein substrate exhibits a different fluorescent property in the phosphorylated state than in the un-phosphorylated state. The protein substrate is then excited and the amount of a fluorescent property that differs in the un-phosphorylated state and phosphorylated state is measured, whereby an amount that is consistent with the presence of the protein substrate in its phosphorylated state indicates the presence of PK activity. The method and products can be used in drug screening. They can be used for screening for compounds which affect cellular events, including receptor-ligand binding, protein-protein interactions or kinase activation, which signal to the target kinase.

XX SQ Sequence 238 AA;

Query Match 98.9%; Score 1262; DB 19; Length 238;
 Best Local Similarity 98.7%; Pred. No. 3.4e-124;

CC properties and have the ability to respond to ion concentrations via a
CC change in fluorescent characteristics. They are useful as markers for
CC gene expression, tracers of cell lineage or as fusion tags to monitor
CC protein localisation within living cells. The engineered fluorescent
CC proteins are particularly useful for coupling engineered fluorescent
CC proteins to antibodies, nucleic acids or other receptors for use in
CC detection assays, e.g. immunoassays or hybridisation assays. They are
CC useful for tracking the movement of proteins in cells or in systems
CC for detecting induction of transcription and for the simultaneous
CC measurement of two or more processes within cells. Proteins of the
CC invention are also useful as fluorescent energy donors or acceptors
CC as well as biosensors for detecting analytes. They are also useful in
CC fluorescence resonance energy transfer (FRET). The crystal structure
CC of the GFP is useful for designing mutants having altered fluorescent
CC characteristics which are particularly useful to identify amino acids
CC whose substitution alters fluorescent properties of the protein. The
CC crystal structure of the GFP is also useful for designing mutants having
CC altered amino binding characteristics which are particularly useful for
CC identifying amino acids whose substitution alters the specificity and
CC affinity of the binding site to various analytes, and for modifying amino
CC binding and therefore the concentration of the analyte the protein
CC sequence is Acquired victoria green fluorescent protein
XX
SQ Sequence 238 AA:
Query Match 38.9% Score 1262. UP 24. Length 238
Best Local Similarity 38.7% E-Val. 3.4e-144
Matches 255; Conservative 21 Mismatches 21 Indels 31 Gaps 17
QY 1 MSKGELEFGVPELIVEIDGAVNHNKFSVSTENEDDAVYKGLKKEKTCGGRKLVNVLIS (1
DB 1 MSKGELEFGVPELIVEIDGAVNHNKFSVSTENEDDAVYKGLKKEKTCGGRKLVNVLIS (1
QY 61 VTLLVGVQCTSRYPDHKKRDEKKSAPFQGVYKERTIFENMGNKKCRARVRFEDVAV 129
DB 61 VTFSYGVQCTSRYPDHKKRDEKKSAPFQGVYKERTIFENMGNKKCRARVRFEDVAV 120
QY 121 NRLEKLGIDKEDQNTIARKEVYNVNHVYIMAKKRNIEKVNFKPENIEFGVNLAD 189
DB 121 NRLEKLGIDKEDQNTIARKEVYNVNHVYIMAKKRNIEKVNFKPENIEFGVNLAD 180
QY 131 HYQGNIPFGGQGVLPDHNHSTQSAASKQDNKRNKDEEVLEFVTAAGTTHGMDLAK 238
DB 131 HYQGNIPFGGQGVLPDHNHSTQSAASKQDNKRNKDEEVLEFVTAAGTTHGMDLAK 238

Search completed: July 16, 2003, 16:10:56
Job time : 46 secs

4/23/01

Sequence version 1.1
Copyright (c) 1994-2003 GapSoft Inc.

CM protein - protein search using sw model

Run on: July 16, 2003, 14:00:47 : Search time 17 seconds
(without alignment)
394,734 Million cell updates/100

Title: US-09-967-301-2

Perfect score: 1275

Sequence: 1 MSKGEELFNVPIVLVDG.....VILDSVAAGLHGMDEYK 436

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 508470 seqs, 133250420 residues

Total number of hits satisfying chosen parameters: 48470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 04

Listing first 45 summaries

Database :

A_Geneseq_101002.*

- 1: /SID52/qcdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 2: /SID52/qcdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/qcdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/qcdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 5: /SID52/qcdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/qcdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 7: /SID52/qcdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/qcdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/qcdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/qcdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/qcdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/qcdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/qcdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/qcdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/qcdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/qcdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/qcdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/qcdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/qcdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/qcdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/qcdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/qcdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/qcdata/geneseq/geneseq-emb1/AA2002.DAT.*

pred. No. is the number of results predicted by BLAST to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1275	100.0	238	17	AAW05304 Green fluorescent
2	1275	100.0	238	18	AAW024232 Acqueora victoria
3	1275	100.0	238	19	AAW767471 A. victoria green
4	1275	100.0	238	19	AAW761305 A. victoria green
5	1275	100.0	238	19	AAW40479 A. victoria green
6	1275	100.0	238	19	AAW73352 Wild-type green fl.
7	1275	100.0	238	23	AAE16034 Acqueora victoria
8	1274	99.9	238	19	AAW65084 A. victoria green
9	1272	99.8	238	19	AAW761305 A. victoria green
10	1272	99.8	238	19	AAW76110 A. victoria green

11	1272	99.8	238	14	AAW76111 A. victoria green
12	1272	99.8	238	19	AAW50078 A. victoria green
13	1272	99.8	238	19	AAW50079 A. victoria green
14	1272	99.8	238	23	AAE16042 Acqueora victoria
15	1272	99.8	238	23	AAE16043 Acqueora victoria
16	1272	99.8	501	18	AAW18799 GFP variants 8457
17	1271	99.7	238	17	AAW5312 Green fluorescent
18	1271	99.7	238	18	AAW36057 Green fluorescent
19	1271	99.7	238	18	AAW1295 Acqueora victoria
20	1271	99.7	238	19	AAW76113 A. victoria green
21	1271	99.7	238	19	AAW50081 A. victoria green
22	1271	99.7	238	19	AAW50083 A. victoria green
23	1271	99.7	238	20	AAV43243 Green fluorescent
24	1271	99.7	238	20	AAW88484 Green fluorescent
25	1271	99.7	238	21	AAW16116 Acqueora victoria
26	1271	99.7	238	21	AAW03411 Acqueora victoria
27	1271	99.7	238	21	AAW15661 Humanised GFP, HC
28	1271	99.7	238	21	AAW01602 Acqueora victoria
29	1271	99.7	238	21	AAV99804 Acqueora victoria
30	1271	99.7	238	21	AAV99835 Acqueora victoria
31	1271	99.7	238	21	AAV99723 Acqueora victoria
32	1271	99.7	238	21	AAV99732 Acqueora victoria
33	1271	99.7	238	21	AAV70365 Humanised green fl
34	1271	99.7	238	22	AAW73554 Green fluorescent
35	1271	99.7	238	22	AAW85699 A. victoria green
36	1271	99.7	238	23	AAW66199 A. victoria green
37	1271	99.7	238	23	AAE14600 Acqueora victoria
38	1271	99.7	238	23	AAE16041 Acqueora victoria
39	1271	99.7	238	23	AAE16056 Acqueora victoria
40	1271	99.7	239	21	AAV69283 Amino acid sequenc
41	1270	99.6	238	17	AAW05306 Green fluorescent
42	1270	99.6	238	19	AAW59822 Green fluorescent
43	1270	99.6	238	19	AAW50086 A. victoria green
44	1270	99.6	238	19	AAW50082 A. victoria green
45	1270	99.6	238	22	AAW73556 Green fluorescent

ALIGNMENTS

RESULT 1
AAW05304
TC AAW05304 Standard; Protein: 238 AA.

XX AAW05304:

XX AC AAW05304:

XX DT 02-APR-1997 (first entry)

XX DE Green fluorescent protein.

XX KW Green fluorescent protein; GFP; jellyfish; Acqueora victoria; sea pansy;

XX KW Retilia reniformis; differential gene expression; protein localisation;

XX KW gene expression; tracking; fluorescence.

XX CS Acqueora victoria.

XX FN WC952380-A1.

XX FC 06-AUG-1996.

XX PF 13-NOV-1995; 95WC-0514642.

XX PR 10-NOV-1994; 94CS-0337315.

XX (RSCC) UNIV CALIFORNIA.

XX Helm E. Tsien RV;

XX WPI; 1996-171470/17.

XX N-PSDB; AAT39644.

XX New modified Acqueora green fluorescent polypeptide(s) having

PT amino acid changes to provide prods. which exhibit different

21 excitation and emission spectra
 XX
 PS Disclosure: Page 20-21: 9pp: English.
 XX
 CC This sequence represents the wild type Aequorea victoria (Jellyfish)
 CC green fluorescent protein (GFP). The fluorescence of this protein is
 CC generated by cyclisation and oxidation of the Ser-Tyr-Gly sequence at
 CC positions 65-67. Aequorea GFP has two absorbance peaks, as opposed to
 CC the one absorbance peak seen in the related GFP from the sea pansy
 CC (Renilla reniformis). This protein is modified to provide the GFPs of
 CC the invention. The modifications were created by substituting the CDNA
 CC encoding this sequence to site directed mutagenesis using mutagenic PCR
 CC primers, or hydroxylamine treatment. The GFPs of the invention are
 CC modified to lead to the formation of products with markedly different
 CC excitation and emission spectra. Visibly distinct colours, and increased
 CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.
 XX
 50 Sequence: 238 AA:
 Query Match: 100.0%; Score 1275; DB: 18; Length: 238;
 Best Local Similarity: 100.0%; Pred. No. 4.6e-124;
 Matches: 238; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;
 QY 1 MSKGEELFTGVPTLVDELGDVNGHKFVSQSGEGDATYKTLKPLDTTKEFVEMTL 60
 DB 1 MSKGEELFTGVPTLVDELGDVNGHKFVSQSGEGDATYKTLKPLDTTKEFVEMTL 60
 QY 61 VTFSYGVQCFESRYPDINKKHDFKSAPEGYVQERTIFFKDGNGYKIPRAENKFGDILV 120
 DB 61 VTFSYGVQCFESRYPDINKKHDFKSAPEGYVQERTIFFKDGNGYKIPRAENKFGDILV 120
 QY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHMVLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHMVLEFVTAAGTTHGMDELYK 238
 RESULT 2
 AAW24212 standard: Protein: 238 AA.
 XX
 AC AAW24212:
 XX
 DT 02-FEB-1998 (first entry)
 DE Aequorea victoria green fluorescent protein (GFP).
 XX
 KW North West Pacific jellyfish; green fluorescent protein; GFP; S60C;
 KW fluorescent resonance energy transfer; FRET; enzymatic assay;
 KW enzyme inhibitor; enzyme promoter; tandem fluorescent protein construct;
 KW blue fluorescent protein; protease cleavage site.
 XX
 OS Aequorea victoria.
 XX
 PN W09726261-A1.
 PD 07-AUG-1997.
 XX
 PF 31-JAN-1997; 97WO-0501457.
 PR 31-JAN-1996; 96US-0594575.
 XX
 PA (AUCO-) AUCORA BIOSCIENCES CORP.
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Cubitt A. Helm R. Tsien RY.
 XX

DR WP2: 1997-4-26:15/37.
 DR N-PSDB: AAT77843.
 XX
 PT Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker,
 PT useful in enzymatic assays
 XX
 58 Disclosure: Fig 1: 8pp: English.
 XX
 CC This sequence represents Aequorea victoria (North West Pacific jellyfish)
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S60C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of
 CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-15 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vitro and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.
 XX
 50 Sequence: 238 AA:
 Query Match: 100.0%; Score 1275; DB: 18; Length: 238;
 Best Local Similarity: 100.0%; Pred. No. 4.6e-124;
 Matches: 238; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;
 QY 1 MSKGEELFTGVPTLVDELGDVNGHKFVSQSGEGDATYKTLKPLDTTKEFVEMTL 60
 DB 1 MSKGEELFTGVPTLVDELGDVNGHKFVSQSGEGDATYKTLKPLDTTKEFVEMTL 60
 QY 61 VTFSYGVQCFESRYPDINKKHDFKSAPEGYVQERTIFFKDGNGYKIPRAENKFGDILV 120
 DB 61 VTFSYGVQCFESRYPDINKKHDFKSAPEGYVQERTIFFKDGNGYKIPRAENKFGDILV 120
 QY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
 DB 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRNTHNSVOLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHMVLEFVTAAGTTHGMDELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTQSALSKEPNKROHMVLEFVTAAGTTHGMDELYK 238
 RESULT 3
 AAW76371 standard: Protein: 238 AA.
 XX
 AC AAW76371:
 XX
 DT 03-DEC-1998 (first entry)
 XX
 DE A. victoria green fluorescent protein.
 XX
 KW Green fluorescent protein; GFP; jellyfish; detection; mutant; activity;
 KW distribution.
 XX
 OS Aequorea victoria.
 XX
 PN JP10234382-A.
 XX
 FC 08-SEP-1998.
 XX
 DE 27-FEB-1997; 97JP-0062370.
 XX
 FF 27-FEB-1997; 97JP-0062370.
 XX
 PA (CINA-) DINABEKU KENKYUSHO KK.
 XX

DR WP: 1998-050564/46.

DE X-PSCH: AAW61751.

XX Green fluorescent protein useful for determining promoter and
 PT targeting activity in a cell

PS Disclosures: Page 6-7: 11pt Japanese.

XX This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from *Aequorea victoria*. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 7 times
 CC higher than the conventional fluorescent proteins. The nucleic acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested.
 CC comprising steps of introducing the above vector to a host cell and
 CC detecting the fluorescence emitted from the cell. They can additionally
 CC be used for detecting the targeting activity in the cell of an amino
 CC acid sequence to be tested in which the above fluorescent protein is
 CC introduced to a cell and the distribution of the fluorescent protein in
 CC the cell is observed.

XX Sequence 238 AA;

Query Match: 100.0%; Score 1275; 58 19; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVELDGVNKHKFSVSGEGSDATYGLKLFCTGKLPVNPITL 60
 DB 1 MSKGEELFTGVVPIVELDGVNKHKFSVSGEGSDATYGLKLFCTGKLPVNPITL 60
 QY 61 VTFTSVGVCFSSYPDHMKRHDFPKSAMPEGVVQERTIFFKCDGNKTRAEVKEGDTLV 120
 DB 61 VTFTSVGVCFSSYPDHMKRHDFPKSAMPEGVVQERTIFFKCDGNKTRAEVKEGDTLV 120
 QY 121 NRLEKLGIDKEKGNLTGKLEYNYNHNVYINADKQNSIKVNFKIPINIGSVCLAD 180
 DB 121 NRLEKLGIDKEKGNLTGKLEYNYNHNVYINADKQNSIKVNFKIPINIGSVCLAD 180
 QY 181 HYQONTPIGPGVLLPUNYLSVSTOSALSKEPKRHMVLEFVTAAGITRHMDELYK 238
 DB 181 HYQONTPIGPGVLLPUNYLSVSTOSALSKEPKRHMVLEFVTAAGITRHMDELYK 238

RESULT 4

AAW76105

ID AAW76105 standard; protein: 238 AA.

XX AAW76105;

XX AAW76105;

DT 18-NOV-1998 (first entry)

DE A. victoria green fluorescent protein.

XX Green fluorescent protein: GFP; mutant: jellyfish; excitatory
 KW chromophore.

OS *Aequorea victoria*.

PR US9254387-A.

PD 08-SEP-1998.

PE 31-JAN-1997; 97US-0791332.

PR 01-FEB-1996; 96US-0610560.

PR 31-JAN-1997; 97US-0791332.

FA (STRO) UNIV IZLAND STANFORD JUNIOR.

XX Cermack BP, Falkow S, Valdivia RH;

XX WP: 1998-050564/43.

XX DNA encoding mutant green fluorescent pigment proteins - with
 PT treated fluorescence intensity than wild-type proteins. useful for
 PT studying gene expression and protein localisation

PS Disclosures: Page 6-7: 15pt; English.

XX This sequence represents the green fluorescent protein (GFP) from
 CC *Aequorea victoria*. This sequence is used in a method to discover GFP's
 CC mutated in the chromophore region that fluoresce more brightly than
 CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 CC method for analyzing a cell containing the GFP. The GFP can also be used
 CC to a protein and used to identify the intracellular localisation of a
 CC protein of interest. A regulatory element could be operatively connected
 CC to a coding portion encoding a mutant GFP and exposed to an environmental
 CC stimulus. The fluorescence signal from the cell then measures
 CC the effect of the stimulus on the regulatory element.
 CC Note: This sequence does not appear in the specification but has
 CC been retrieved from the EMBL record contained in W05304.

XX Sequence 238 AA;

Query Match: 100.0%; Score 1275; 58 19; Length 238;
 Best Local Similarity 100.0%; Pred. No. 4.6e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVELDGVNKHKFSVSGEGSDATYGLKLFCTGKLPVNPITL 60
 DB 1 MSKGEELFTGVVPIVELDGVNKHKFSVSGEGSDATYGLKLFCTGKLPVNPITL 60
 QY 61 VTFTSVGVCFSSYPDHMKRHDFPKSAMPEGVVQERTIFFKCDGNKTRAEVKEGDTLV 120
 DB 61 VTFTSVGVCFSSYPDHMKRHDFPKSAMPEGVVQERTIFFKCDGNKTRAEVKEGDTLV 120
 QY 121 NRLEKLGIDKEKGNLTGKLEYNYNHNVYINADKQNSIKVNFKIPINIGSVCLAD 180
 DB 121 NRLEKLGIDKEKGNLTGKLEYNYNHNVYINADKQNSIKVNFKIPINIGSVCLAD 180
 QY 181 HYQONTPIGPGVLLPUNYLSVSTOSALSKEPKRHMVLEFVTAAGITRHMDELYK 238
 DB 181 HYQONTPIGPGVLLPUNYLSVSTOSALSKEPKRHMVLEFVTAAGITRHMDELYK 238

RESULT 5

AAW40479

ID AAW40479 standard; protein: 236 AA.

XX AAW40479;

XX AAW40479;

DT 15-SEP-1998 (first entry)

DE A. victoria green fluorescent protein.

XX Green fluorescent protein: protein kinase; phosphate donor;
 KW phosphorylation; drug screening; receptor-ligand binding; signalling;
 KW protein-protein interaction; kinase activation.

OS *Aequorea victoria*.

PR Key Local No./unifications

PD 1.236 /no/0= "partial sequence"

PE WC9502571-A1.

PR 22-JAN-1998.

PD 16-JUL-1997; 97WO-US2410.

PR 16-JUL-1996; 96US-0680877.

PR 16-JUL-1996; 96US-0679865.

PR 16-JUL-1996; 96US-0680876.

(REGC) UNIV CALIFORNIA.

XX Qubitt Ab, Tsien KY;

XX WPI: 1998-110416/10.

DR N-PSDB: AAV11125.

XX Assays for protein kinase and modulators - using a fluorescent

PT protein substrate which exhibits a different fluorescent property in

PT the phosphorylated and unphosphorylated state

XX

PS Disclosure: Fig 3: 55pp; English.

XX This sequence represents a green fluorescent protein from Aequorea victoria which is used in a method to determine if a sample contains protein kinase (PK) activity. The method involves contacting the sample with a phosphate donor (PD) and a fluorescent protein (FP) substrate for a PK, the protein substrate comprising a fluorophore and a phosphorylation site for a PK, where the protein substrate exhibits a different fluorescent property in the phosphorylated state than in the unphosphorylated state. The protein substrate is then excited and the amount of a fluorescent property that differs in the unphosphorylated state and phosphorylated state is measured, whereby an amount that is consistent with the presence of the protein substrate in its phosphorylated state indicates the presence of PK activity. The method and products can be used in drug screening. They can be used for screening for compounds which affect cellular events, including receptor-ligand binding, protein-protein interactions or kinase activation, which signal to the target kinase.

XX Sequence 238 AA:

SQ Query Match 100.0%; Score 1275; DB 10; Length 238;

Best Local Similarity 100.0%; Pred. No. 4, 60-124;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGGATYGLKILKFKICTGKLPVWPTL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGGATYGLKILKFKICTGKLPVWPTL 60
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKSAPEGYVOERTIFFKDGNTKTRAFVKEGDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 61 VTFESYGVOCFSRYPDHMKRPHDFKSAPEGYVOERTIFFKDGNTKTRAFVKEGDTLV 120
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRLEKGIQFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGSVQLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 121 NRLEKGIQFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGSVQLAD 180
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238
 DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6

AAB73552

ID AAB73552 standard; Protein: 238 AA.

XX

AC AAB73552;

XX

XX 07-AUG-2003; (first entry)

XX

DE Wild-type green fluorescent protein (wtGFP).

XX

KW Phenotype selection; non-selectable; fusion protein; stable expression;

KW selectable marker; antibiotic resistance gene; Escherichia coli;

KW green fluorescent protein; GFP; wtGFP; pGFP.

XX

CS Aequorea victoria.

XX

PN WC200120225-A1.

XX

XX 26-APR-2003.

XX

29-MAR-2000; 2000WO-0508477.

XX

21-OCT-1999; 98US-0160461.

XX

22-FEB-2000; 2000US-0510097.

XX

(PANG-) PANORAMA RES INC.

XX

Salint RF;

XX

WPI: 2001-202162/29.

XX

N-PSDB: AAB20245.

XX

Obtaining cells expressing mutant protein, comprises selecting from

XX cells transformed with library of mutagenized protein coding sequences

XX joined to selector protein, which confers growth under selective

XX conditions

XX

Example 2; Page 1; 55pp; English.

XX

The invention relates to methods whereby variants of proteins which do

XX not confer selectable phenotypes can be selected for stable expression

XX in host cells (especially Escherichia coli). The methods can be used to

XX obtain mutants of a desired protein optimised for expression in the host

XX cells, for obtaining a mutant protein with enhanced stability relative to

XX the wild-type protein, and for identifying peptides that stabilise an

XX unstable protein. The methods all involve expressing the protein of

XX interest as a fusion with a protein that can confer a selectable

XX phenotype, such as an antibiotic resistance protein. The transformed

XX host cells are then grown under selective pressure (e.g., presence of

XX antibiotic). The cells able to grow under such conditions are those

XX which contain fusion proteins which are optimised for expression or

XX which are more stable, as this property will also correlate with an

XX increased amount or rate of synthesis of the selectable marker. The

XX invention also discloses mutants of green fluorescent protein (GFP)

XX selected for increased stability using the method of the invention.

XX The present sequence represents wild-type GFP (wtGFP) as encoded by

XX plasmid pGFP (Genbank G17597), which was used in an exemplification

XX of the invention.

XX The present sequence is not shown in the specification, but was

XX derived from GenBank accession number U7997.

XX

SQ Sequence 238 AA:

XX

Query Match 100.0%; Score 1275; DB 22; Length 238;

Best Local Similarity 100.0%; Pred. No. 4, 60-124;

Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGGATYGLKILKFKICTGKLPVWPTL 60

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGGATYGLKILKFKICTGKLPVWPTL 60

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 VTFESYGVOCFSRYPDHMKRPHDFKSAPEGYVOERTIFFKDGNTKTRAFVKEGDTLV 120

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 VTFESYGVOCFSRYPDHMKRPHDFKSAPEGYVOERTIFFKDGNTKTRAFVKEGDTLV 120

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 NRLEKGIQFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGSVQLAD 180

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 NRLEKGIQFKEDGNILGHKLEYNYNHNVYIMADKQKNGIKVNFKIRINIEFGSVQLAD 180

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 HYQONTPIGDGPVLLPNHNYLSTQSALSKDPNKRKHNVILLEFVTAGITHGMDELYK 238

DE :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7

AAB73538

ID AAB73538 standard; Protein: 238 AA.

XX

XX AAB73538;

XX

XX 26-MAR-2002 (first entry)

XX

XX Aequorea victoria green fluorescent protein (GFP).

XX Green fluorescent protein; cell lineage tracer; protein localization
 KW GFP; fusion tag; gene expression marker; fluorescent energy acceptor;
 KW immunassay; hybridization assay; fluorescent energy donor; fusion
 KW FRET; fluorescence resonance energy transfer.
 XX Aequorea victoria.
 XX W020050147-A2.
 XX 29-NOV 2001.
 XX 17-MAY-2001; 2001WO-0576147.
 XX 19-MAY-2003; 2003US-0575847.
 XX (UYCR-) UNIV OREGON STATE.
 XX Wachter R. Remington SJ;
 FI WPI; 2002-083084/11.
 DR N-PSDB; AAD26197.
 XX New long wavelength engineered fluorescent proteins; useful as markers
 PT for gene expression, tracers of cell lineage or as fusion tags to
 PT monitor protein localization, or in detection assays, e.g. immunoassays
 PT or hybridization assays -
 XX Claim 27; Fig 3; 181pp; English.
 XX The patent discloses functional engineered fluorescent protein and its
 CC corresponding polynucleotide. The amino acid sequence of the engineered
 CC protein is identical to Aequorea green fluorescent protein (GFP). The
 CC engineered fluorescent proteins of the invention have varied fluorescent
 CC properties and have the ability to respond to ion concentrations via a
 CC change in fluorescent characteristics. They are useful as markers for
 CC gene expression, tracers of cell lineage or as fusion tags to monitor
 CC protein localization within living cells. The engineered fluorescent
 CC proteins are particularly useful for coupling engineered fluorescent
 CC proteins to antibodies, nucleic acids or other receptors for use in
 CC detection assays, e.g. immunoassays or hybridization assays. They are
 CC useful for tracking the movement of proteins in cells or in systems
 CC for detecting induction of transcription and for the simultaneous
 CC measurement of two or more processes within cells. Proteins of the
 CC invention are also useful as fluorescent energy donors or acceptors
 CC as well as biosensors for detecting actions. They are also useful as
 CC fluorescence resonance energy transfer (FRET) tags. The crystal structure
 CC of the GFP is useful for designing mutants having altered fluorescent
 CC characteristics which are particularly useful to identify amino acids
 CC whose substitution alters fluorescent properties of the protein. The
 CC crystal structure of the GFP is also useful for designing mutants having
 CC altered amino binding characteristics which are particularly useful for
 CC identifying amino acids whose substitution affects the specificity and
 CC affinity of the binding site to various amino acids and for mutated amino
 CC binding and therefore the concentration of the amino. The present
 CC sequence is Aequorea victoria green fluorescent protein.
 XX
 SO Sequence 238 AA:
 Query Match: 100.00; Score 1275; DP 23; Length 238;
 Best Local Similarity 100.00; Pred. No. 4, 5e-124;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGELETVGVVPIIVEDGDGVNKHKFSVSGESDANYGKLTLEKTTGKIVPPEIL 60
 DB 1 MSKGELETVGVVPIIVEDGDGVNKHKFSVSGESDANYGKLTLEKTTGKIVPPEIL 60
 QY 61 VTTSYGVQCISRYPDHKKRDFKSAFPEGYVQERTFFKDDGNKTKRAEVKPEGDIIV 120
 DB 61 VTTSYGVQCISRYPDHKKRDFKSAFPEGYVQERTFFKDDGNKTKRAEVKPEGDIIV 120
 QY 121 NKIEIKGIDFKEDSNLGHKLEYNNSHNYIMADKKNGIKVNFIRHKLEKSVGLAD 180
 DB 121 NKIEIKGIDFKEDSNLGHKLEYNNSHNYIMADKKNGIKVNFIRHKLEKSVGLAD 180

DB 121 NKIEIKGIDFKEDSNLGHKLEYNNSHNYIMADKKNGIKVNFIRHKLEKSVGLAD 180
 QY 161 HYQONTFADGGVGLFADNHYSLTQSAKSKDPEKSDHMYVLSEFVTAAGTTHGMDELYK 248
 DB 161 HYQONTFADGGVGLFADNHYSLTQSAKSKDPEKSDHMYVLSEFVTAAGTTHGMDELYK 248
 RESULT 8
 AAW65084
 ID AAW65084 standard; Protein; 234 AA.
 XX
 AC AAW65084;
 XX
 DI 15-SEP-1998 (first entry)
 XX
 DE A. victoria green fluorescent protein variant E171K.
 XX
 KW Green fluorescent protein; protein kinase; phosphate donor;
 KW phosphorylation; drug screening; receptor-ligand binding; signaling;
 KW protein-protein interaction; kinase activation.
 XX Aequorea victoria.
 CS Synthetic.
 PH Key Location/Qualifiers
 FT Misc-difference 172 /label= E171K
 FT /note= wild type Gln is replaced with Lys. The
 FT specification states the mutation occurs at
 FT position 171 however it is more likely to be at
 FT position 172.
 XX W02002571-A1.
 XX 22-JAN-1998.
 XX 16-JUL-1996; 96US-0680477.
 XX 16-JUL-1996; 96US-0678465.
 XX 16-JUL-1996; 96US-0680476.
 XX (REGO) UNIV CALIFORNIA.
 PI Cubitt AB, Tsien RY;
 DR WPI; 1998-110616/10.
 XX Assays for protein kinase and modulators - using a fluorescent
 FT protein substrate which exhibits a different fluorescent property in
 FT the phosphorylated and un-phosphorylated state
 XX
 PS Claim 35; Page 1; 65pp; English.
 XX This sequence represents a green fluorescent protein variant from
 CC Aequorea victoria in which a wild-type Gln residue at position 172 is
 CC replaced with a Lys. The specification states that this mutation occurs
 CC at position 171 however the wild type Gln residue is located at position
 CC 172. This variant is used in a method to determine if a sample contains
 CC protein kinase (PK) activity. The method involves contacting the sample
 CC with a phosphate donor (DP) and a fluorescent protein (FP) substrate for
 CC a PK, the protein substrate comprising a FP moiety and a phosphorylation
 CC site for a PK, where the protein substrate exhibits a different
 CC fluorescent property in the phosphorylated state than in the
 CC un-phosphorylated state. The protein substrate is then excited and the
 CC amount of a fluorescent property that differs in the un-phosphorylated
 CC state and phosphorylated state is measured, whereby an amount that is
 CC consistent with the presence of the protein substrate in its
 CC phosphorylated state indicates the presence of PK activity. The method
 CC and products can be used in drug screening. They can be used for
 CC screening for compounds which affect cellular events, including
 CC receptor-ligand binding, protein-protein interactions or kinase
 CC activation, which signal to the target kinase.

PS Claim 6; Page 7; 15pp; English.

XX This sequence represents a mutant green fluorescent protein (GFP) from

XX Acetivorea victoria in which a Ser residue at position 65 in the wild

CC type protein is replaced by an Ala. These mutations occur in the

CC chromophore region. GFP mutants are used in a method to discover GFP's

CC mutated in the chromophore region that fluoresce more brightly than

CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a

CC method for analyzing a cell containing the GFP. The GFP can also be

CC fused to a protein and used to identify the intracellular localisation of

CC a protein of interest. A regulatory element could be operatively

CC connected to a coding portion encoding a mutant GFP and exposed to an

CC environmental stimulus. The fluorescence signal from the cell

CC then measures the effect of the stimulus on the regulatory element.

CC NOTE: This sequence does not appear in the specification but has been

CC constructed from the wild-type GFP protein represented in AAW76116

XX

SQ Sequence 238 AA:

Query Match 99.8% Score 1272; DB 19; Length 238;

Best Local Similarity 99.6%; Pred. No. 9,36-124;

Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLELDGVNCHKFSVSGEGSDATYKGLTKFKICTGKLPVPRPTL 60

DE 1 MSKGEELFTGVVPIVLELDGVNCHKFSVSGEGSDATYKGLTKFKICTGKLPVPRPTL 60

QY 61 VTFESYGVQCFSPRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEETLV 120

DE 61 VTFESYGVQCFSPRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEETLV 120

QY 121 NRLEKIGIDFKEDGNLGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNTEGSSVQLAD 180

DE 121 NRLEKIGIDFKEDGNLGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNTEGSSVQLAD 180

QY 181 HYQONTPIGDSPIVLPDKNHLYSTQSALSKDPNEKRDHNVLEFVTAAGITHGMDELYK 238

DE 181 HYQONTPIGDSPIVLPDKNHLYSTQSALSKDPNEKRDHNVLEFVTAAGITHGMDELYK 238

RESULT 11

AAW76111

ID AAW76111 standard; protein: 238 AA.

XX

XX AAW76111:

XX

XX 18-NOV-1998 (first entry)

XX

XX A. victoria green fluorescent protein mutant: vawc.

XX

XX Green fluorescent protein: GFP; mutant: yellowish excitation;

XX chromophore.

XX

XX Acetivorea victoria.

XX

XX Synthetic.

XX

XX Key Location/Qualifiers

FT Misc-difference 65

FT /label= Val

FT /note= "wild-type Val is replaced by Ala"

XX

XX US5824367-A.

XX

XX 08-SEP-1998.

XX

XX 31-JAN-1997; 97US-0791332.

XX

XX 01-FEB-1996; 96US-0010760.

XX

XX 31-JAN-1997; 97US-0791332.

XX

XX (STRD) UNIV LELAND STANFORD JUNIOR.

XX

XX Cormack BP, Faikow S, Valdivia RH.

XX

XX WP1: 1598 505545/43.

XX

XX DNA encoding mutant green fluorescent pigment proteins - with

XX greatest fluorescence intensity than wild-type proteins, useful for

XX studying gene expression and protein localisation

XX

XX Claim 7; Page 7; 15pp; English.

XX This sequence represents a mutant green fluorescent protein (GFP) from

XX Acetivorea victoria in which a Val residue at position 68 in the wild

CC type protein is replaced by a Leu. These mutations occur in the

CC chromophore region. GFP mutants are used in a method to discover GFP's

CC mutated in the chromophore region that fluoresce more brightly than

CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a

CC method for analyzing a cell containing the GFP. The GFP can also be

CC fused to a protein and used to identify the intracellular localisation of

CC a protein of interest. A regulatory element could be operatively

CC connected to a coding portion encoding a mutant GFP and exposed to an

CC environmental stimulus. The fluorescence signal from the cell

CC then measures the effect of the stimulus on the regulatory element.

CC NOTE: This sequence does not appear in the specification but has been

CC constructed from the wild-type GFP protein represented in AAW76116

XX

SQ Sequence 238 AA:

Query Match 99.8% Score 1272; DB 19; Length 238;

Best Local Similarity 99.6%; Pred. No. 9,36-124;

Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLELDGVNCHKFSVSGEGSDATYKGLTKFKICTGKLPVPRPTL 60

DE 1 MSKGEELFTGVVPIVLELDGVNCHKFSVSGEGSDATYKGLTKFKICTGKLPVPRPTL 60

QY 61 VTFESYGVQCFSPRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEETLV 120

DE 61 VTFESYGVQCFSPRYPDHMKRHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKFEETLV 120

QY 121 NRLEKIGIDFKEDGNLGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNTEGSSVQLAD 180

DE 121 NRLEKIGIDFKEDGNLGHKLEYNNSHNVYIMADQKNGIKVNFKIRHNTEGSSVQLAD 180

QY 181 HYQONTPIGDSPIVLPDKNHLYSTQSALSKDPNEKRDHNVLEFVTAAGITHGMDELYK 238

DE 181 HYQONTPIGDSPIVLPDKNHLYSTQSALSKDPNEKRDHNVLEFVTAAGITHGMDELYK 238

RESULT 12

AAW65278

ID AAW65278 standard; protein: 238 AA.

XX

XX AAW65278:

XX

XX 18-SEP-1996 (first entry)

XX

XX A. victoria green fluorescent protein variant S65T.

XX

XX Green fluorescent protein: protein kinase; phosphate donor;

XX phosphorylation: drug screening; receptor-ligand binding; signalling;

XX protein-protein interaction; kinase activation.

XX

XX Acetivorea victoria.

XX

XX Synthetic.

XX

XX Key Location/Qualifiers

FT Misc-difference 65

FT /label= S65T

FT /note= "wild type Ser is replaced with Thr"

XX

XX W09802571-A1.

XX

XX 22-JAN-1998.

XX

PF 16-JUL-1997; 97WO-US12410.
 XX
 PR 16-JUL-1996; 96US-0680877.
 PR 16-JUL-1996; 96US-0679865.
 PR 16-JUL-1996; 96US-0680876.
 XX
 VA (REGC) UNIV CALIFORNIA.
 XX
 PI Cubitt AB, Tsien KY;
 XX WPI: 1998-110616/10.
 XX
 XX Assays for protein kinase and modulators - using a fluorescent
 PI protein substrate which exhibits a different fluorescent property in
 PT the phosphorylated and un-phosphorylated state
 XX
 PS Claim 27: Page -: 65pp; English.
 XX
 CC This sequence represents a green fluorescent protein variant from
 CC Aequorea victoria in which a wild-type Ser residue at position 65 is
 CC replaced with a Thr. This variant is used in a method to determine if a
 CC sample contains protein kinase (PK) activity. The method involves
 CC contacting the sample with a phosphate donor (PD) and a fluorescent
 CC protein (FP) substrate for a PK, the protein substrate comprising a FP
 CC moiety and a phosphorylation site for a PK, where the protein substrate
 CC exhibits a different fluorescent property in the phosphorylated state
 CC than in the un-phosphorylated state. The protein substrate is then
 CC excited and the amount of a fluorescent property that differs in the
 CC un-phosphorylated state and phosphorylated state is measured, whereby an
 CC amount that is consistent with the presence of the protein substrate in
 CC its phosphorylated state indicates the presence of PK activity. The
 CC method and products can be used in drug screening. They can be used for
 CC screening for compounds which affect cellular events, including
 CC receptor-ligand binding, protein-protein interactions or kinase
 CC activation, which signal to the target kinase.
 CC NOTE: This sequence does not appear in the specification and has
 CC been constructed from the wild-type sequence represented in AAK40479.
 XX
 SQ Sequence 238 AA:
 Query Match: 99.9%; Score 1272; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKILVFWPTL 60
 DB 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKILVFWPTL 60
 QY 61 VTTFYGVQCFSRYPDHMKRHDFKFSAMPFGYVQERTIFFKDKNKYKRAEVKFGCTLV 120
 DB 61 VTTFYGVQCFSRYPDHMKRHDFKFSAMPFGYVQERTIFFKDKNKYKRAEVKFGCTLV 120
 QY 121 NRIELKGIDFKEDSNILGHKLEYNINSHNVIYMAQKQNGIKVNFKIRHNLEDSVOLAD 180
 DB 121 NRIELKGIDFKEDSNILGHKLEYNINSHNVIYMAQKQNGIKVNFKIRHNLEDSVOLAD 180
 QY 181 HYQONTPIGDSGVLLPNNHYLSTQSALSQKDPNFKRHMHVLEFVTAAGTHTGMLEYK 238
 DB 181 HYQONTPIGDSGVLLPNNHYLSTQSALSQKDPNFKRHMHVLEFVTAAGTHTGMLEYK 238

Protein-protein interaction: kinase activation.
 Aequorea victoria.
 Synthetic.
 Key Location/Otifiers
 Misc-difference 65 /label= 565A
 /note= "wild type Ser is replaced with Ala"
 WC9802571-A1.
 22-JAN-1998; 97WO-US12410.
 16-JUL-1997; 97WO-US12410.
 16-JUL-1996; 96US-0680877.
 16-JUL-1996; 96US-0679865.
 16-JUL-1996; 96US-0680876.
 (REGC) UNIV CALIFORNIA.
 Cubitt AB, Tsien KY;
 WPI: 1998-110616/10.
 Assays for protein kinase and modulators - using a fluorescent
 protein substrate which exhibits a different fluorescent property in
 the phosphorylated and un-phosphorylated state
 Claim 27: Page -: 65pp; English.
 This sequence represents a green fluorescent protein variant from
 Aequorea victoria in which a wild-type Ser residue at position 65 is
 replaced with a Ala. This variant is used in a method to determine if a
 sample contains protein kinase (PK) activity. The method involves
 contacting the sample with a phosphate donor (PD) and a fluorescent
 protein (FP) substrate for a PK, the protein substrate comprising a FP
 moiety and a phosphorylation site for a PK, where the protein substrate
 exhibits a different fluorescent property in the phosphorylated state
 than in the un-phosphorylated state. The protein substrate is then
 excited and the amount of a fluorescent property that differs in the
 un-phosphorylated state and phosphorylated state is measured, whereby an
 amount that is consistent with the presence of the protein substrate in
 its phosphorylated state indicates the presence of PK activity. The
 method and products can be used in drug screening. They can be used for
 screening for compounds which affect cellular events, including
 receptor-ligand binding, protein-protein interactions or kinase
 activation, which signal to the target kinase.
 NOTE: This sequence does not appear in the specification and has
 been constructed from the wild-type sequence represented in AAK40479.
 SQ Sequence 238 AA:
 Query Match: 99.9%; Score 1272; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 9,3e-124;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKILVFWPTL 60
 DB 1 MSKGEELFTGVVFLVDELGVNCHKFSVSGSGEDATYCKILKFKICTGKILVFWPTL 60
 QY 61 VTTFYGVQCFSRYPDHMKRHDFKFSAMPFGYVQERTIFFKDKNKYKRAEVKFGCTLV 120
 DB 61 VTTFYGVQCFSRYPDHMKRHDFKFSAMPFGYVQERTIFFKDKNKYKRAEVKFGCTLV 120
 QY 121 NRIELKGIDFKEDSNILGHKLEYNINSHNVIYMAQKQNGIKVNFKIRHNLEDSVOLAD 180
 DB 121 NRIELKGIDFKEDSNILGHKLEYNINSHNVIYMAQKQNGIKVNFKIRHNLEDSVOLAD 180
 QY 181 HYQONTPIGDSGVLLPNNHYLSTQSALSQKDPNFKRHMHVLEFVTAAGTHTGMLEYK 238
 DB 181 HYQONTPIGDSGVLLPNNHYLSTQSALSQKDPNFKRHMHVLEFVTAAGTHTGMLEYK 238

change in fluorescent characteristics. They are useful as markers for gene expression, tracers of cell lineage or as fusion tags to monitor protein localisation within living cells. The engineered fluorescent proteins are particularly useful for coupling engineered fluorescent proteins to antibodies, nucleic acids or other receptors for use in detection assays, e.g. immunoassays or hybridisation assays. They are useful for tracking the movement of proteins in cells or in systems for detecting induction of transcription and for the simultaneous measurement of two or more processes within cells. Proteins of the invention are also useful as fluorescent energy donors or acceptors as well as biosensors for detecting anions. They are also useful in fluorescence resonance energy transfer (FRET). The crystal structure of the GFP is useful for designing mutants having altered fluorescent characteristics which are particularly useful to identify amino acids whose substitution alters fluorescent properties of the protein. The crystal structure of the GFP is also useful for designing mutants having altered union binding characteristics which are particularly useful for identifying amino acids whose substitution alters the specificity and affinity of the binding site to various anions, and for monitoring anion binding and therefore the concentration of the anion. The present sequence is Aequorea victoria GFP mutant protein (A6A1).

Note: This sequence is not shown in the specification but is derived from Aequorea victoria wild type green fluorescent protein shown in figure 3 of the specification (A6A1638).

XX Sequence 238 AA:

Query Match 95.2% Score 12721 DP 217 Length 238
 Best Local Similarity 91.6% Prod. No. 9.3e+124
 Matches 237: Conservative 17 Mismatches 17 Indels 0 Gaps 0

07 1 MSKGEELFGVWPIFVELDGVNGKKEFSVSGSDERDAVYGNILKEICTCKIKVWNTIL 60
 20 1 MSKGEELFGVWPIFVELDGVNGKKEFSVSGSDERDAVYGNILKEICTCKIKVWNTIL 60
 07 61 VTTFYSYGVQCFKRYEDHMKRHHFFKSAHPYVQCFTHFKKQKNYKTAAYVAFEGHTLY 120
 06 61 VTTFAYGVQCFKRYEDHMKRHHFFKSAHPYVQCFTHFKKQKNYKTAAYVAFEGHTLY 120
 07 121 NRLEKSLDPRKNNYLDHKLNNKNNNNYDAAKQKNGKAVNFKLQKRTGDSVLAAL 180
 10 121 NRLEKSLDPRKNNYLDHKLNNKNNNNYDAAKQKNGKAVNFKLQKRTGDSVLAAL 180
 07 181 HYQCNTPIDGPPVLLDNNHVLSTQSALSKEINFKSCHHVVLEFVTAAGTQMTFLYK 240
 10 181 HYQCNTPIDGPPVLLDNNHVLSTQSALSKEINFKSCHHVVLEFVTAAGTQMTFLYK 240
 10 181 HYQCNTPIDGPPVLLDNNHVLSTQSALSKEINFKSCHHVVLEFVTAAGTQMTFLYK 240

Search completed: July 16, 2003, 16:10:36
 Job time : 47 secs

XX WPI: 2002-063084/11.

XX New long wavelength engineered fluorescent proteins: useful as markers

XX for gene expression, tracers of cell lineage or as fusion tags to

XX monitor protein localization, or in detection assays, e.g. immunoassays

XX or hybridization assays

XX Example: Page 1: 181pp: English.

XX The patent discloses functional engineered fluorescent protein and its

XX corresponding polynucleotide. The amino acid sequence of the engineered

XX protein is identical to Aequorea green fluorescent protein (GFP). The

XX engineered fluorescent proteins of the invention have varied fluorescent

XX properties and have the ability to respond to ion concentrations via a

XX change in fluorescent characteristics. They are useful as markers for

XX gene expression, tracers of cell lineage or as fusion tags to monitor

XX protein localisation within living cells. The engineered fluorescent

XX proteins are particularly useful for coupling engineered fluorescent

XX proteins to antibodies, nucleic acids or other receptors for use in

XX detection assays, e.g. immunoassays or hybridisation assays. They are

XX useful for tracking the movement of proteins in cells or in systems

XX for detecting induction of transcription and for the simultaneous

XX measurement of two or more processes within cells. Proteins of the

XX invention are also useful as fluorescent energy donors or acceptors

XX as well as biosensors for detecting analytes. They are also useful in

XX fluorescence resonance energy transfer (FRET). The crystal structure

XX of the GFP is useful for designing mutants having altered fluorescent

XX characteristics which are particularly useful to identify amino acids

XX whose substitution alters fluorescent properties of the protein. The

XX crystal structure of the GFP is also useful for designing mutants having

XX altered amino acid binding characteristics which are particularly useful for

XX identifying amino acids whose substitution alters the specificity and

XX affinity of the binding site to various anions, and for monitoring anion

XX binding and therefore the concentration of the anion. The present

XX sequence is Aequorea victoria GFP mutant protein (B2228).

XX Note: This sequence is not shown in the specification but is derived

XX from Aequorea victoria wild type green fluorescent protein shown in

XX figure 3 of the specification (AA76103H).

XX Sequence 238 AA:

Query Match 99.2% Score 1259 DB 19 Length 238

Best Local Similarity 99.2% Pred. No. 1.5e-123

Matches 235: Conservative 0: Mismatches 1: Indels 0: Gaps 0

QY 1 MSKGEELFTSVPLVDELDDVGNHGFVSVEGEGDATYGKLTAKFICTGKLPVWPTL 60

DB 1 MSKGEELFTSVPLVDELDDVGNHGFVSVEGEGDATYGKLTAKFICTGKLPVWPTL 60

QY 61 VTTLTSGVQCFSSRYPDHMKRHDFFKSAPEGYVQERTIFFRDDGNKYTRAEVKFEGMTLV 120

DB 61 VTTLTSGVQCFSSRYPDHMKRHDFFKSAPEGYVQERTIFFRDDGNKYTRAEVKFEGMTLV 120

QY 121 NRLELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKRHNTEGGVGLAD 180

DB 121 NRLELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKRHNTEGGVGLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSIQSALSQSKDNPKRCHNWVLCGFVTAAGTTHGMDELTK 238

DB 181 HYQONTPIGDPVLLPDNHYLSIQSALSQSKDNPKRCHNWVLCGFVTAAGTTHGMDELTK 238

RESULT 2

AAW76106

ID AAW76106 standard; protein: 238 AA.

XX AAW76106:

XX 18-NOV-1998 (first entry)

XX A. victoria green fluorescent protein mutant #64L/565T

KW Great fluorescent protein: GFP; mutant: jellyfish; excitation; chromophore.

XX Aequorea victoria.

XX Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 64 /label= F64L

XX Misc-difference 65 /note= "Wild-type Phe is replaced with Leu"

XX Misc-difference 66 /label= S65T

XX Misc-difference 67 /note= "Wild-type Ser is replaced by Thr"

XX US5804387-A.

XX 08-SEP-1998.

XX 31-JAN-1997: 97US-0791332.

XX 01-FEB-1996: 96US-0010963.

XX 31-JAN-1997: 97US-0791332.

XX (STRD) UNIV LELAND STANFORD CENTER.

XX Corrack BP, Falkow S, Valdivia RH:

XX WPI: 1998-505643/43.

XX DNA encoding mutant green fluorescent protein (GFP) with greater fluorescence intensity than wild-type proteins, useful for studying gene expression and protein localisation

XX Disclosure: Page 1: 15pp: English.

XX This sequence represents a mutant green fluorescent protein (GFP) from Aequorea victoria in which a Phe residue at position 64 in the wild type protein is replaced by a Leu and a Ser at position 65 is replaced by a Thr. These mutations occur in the chromophore region. GFP mutants are used in a method to discover GFP's mutated in the chromophore region that fluoresce more brightly than wild-type GFP upon excitation at 488 nm. These mutants can be used in a method for analyzing a cell containing the GFP. The GFP can also be fused to a protein and used to identify the intracellular localisation of a protein of interest. A regulatory element could be operatively connected to a coding portion encoding a mutant GFP and exposed to an environmental stimulus. The fluorescence signal from the GFP then measures the effect of the stimulus on the regulatory element.

XX Note: This sequence does not appear in the specification but has been constructed from the wild-type GFP protein represented in AAW76106

XX Sequence 238 AA:

Query Match 99.2% Score 1259 DB 19 Length 238

Best Local Similarity 99.2% Pred. No. 1.5e-123

Matches 235: Conservative 1: Mismatches 2: Indels 0: Gaps 0

QY 1 MSKGEELFTSVPLVDELDDVGNHGFVSVEGEGDATYGKLTAKFICTGKLPVWPTL 60

DB 1 MSKGEELFTSVPLVDELDDVGNHGFVSVEGEGDATYGKLTAKFICTGKLPVWPTL 60

QY 61 VTTLTSGVQCFSSRYPDHMKRHDFFKSAPEGYVQERTIFFRDDGNKYTRAEVKFEGMTLV 120

DB 61 VTTLTSGVQCFSSRYPDHMKRHDFFKSAPEGYVQERTIFFRDDGNKYTRAEVKFEGMTLV 120

QY 121 NRLELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKRHNTEGGVGLAD 180

DB 121 NRLELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKVNFKRHNTEGGVGLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSIQSALSQSKDNPKRCHNWVLCGFVTAAGTTHGMDELTK 238

DB 181 HYQONTPIGDPVLLPDNHYLSIQSALSQSKDNPKRCHNWVLCGFVTAAGTTHGMDELTK 238

CC intensities of emission make these products useful in a wide variety of
 CC contexts, such as tracking of differential gene expression and protein
 CC localisation. The mutations can also be created to modify the encoded
 CC GFP so that it only possesses one absorbance peak.

XX Sequence 236 AA:

Query Match 98.6%; Score 1258; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. No. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 60
 DB 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 60
 QY 61 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 DB 61 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 QY 62 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 DB 62 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 QY 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVNFKIRHNIEDGVVQGLAD 180
 DB 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVNFKIRHNIEDGVVQGLAD 180
 QY 181 HYQONTPTGQGVVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 238
 DB 181 HYQONTPTGQGVVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 238

RESULT 5
 AAW24232 standard: protein; 236 AA.

XX AAW24232;
 AC AAW24232;
 DI 02-FEB-1996 (first entry)
 XX Aequorea victoria green fluorescent protein (GFP);
 XX North West Pacific jellyfish; green fluorescent protein (GFP); GFP;
 KW fluorescent resonance energy transfer (FRET); enzymatic assays;
 KW enzyme inhibitor; enzyme promoter; tandem fluorescent protein constructs;
 KW blue fluorescent protein; protease cleavage site.

XX Aequorea victoria;
 XX W09724261-A1;
 XX 07-AUG-1997;
 XX 31-JAN-1997; 97WG-US01457;
 XX 31-JAN-1996; 96US-0594575;
 PA (AJRO-) Aequorea Biosciences Corp.
 XX (REGC) UNIV CALIFORNIA;
 PI Cubitt A., Heim R., Tsien RY;
 DR WP1: 1997-402615/37;
 DR N-PSDB; AA:77843.

XX Tandem fluorescent protein constructs - have donor and acceptor
 PT moieties exhibiting fluorescent linked via cleavable peptide linker.
 PI useful in enzymatic assays

XX Disclosure: Fig 1; 80pp; English.
 XX This sequence represents Aequorea victoria (North West Pacific jellyfish);
 CC green fluorescent protein (GFP). The GFP protein is used in a novel
 CC tandem fluorescent protein construct. The construct comprises a donor
 CC (e.g. GFP protein) and an acceptor (e.g. S65C, a variant of GFP or
 CC blue fluorescent protein) fluorescent protein moiety, and a linker
 CC coupling them. Preferably, the donor is positioned at the N-terminus of

CC the polypeptide relative to the acceptor. The linker moiety is a peptide
 CC 5-50 amino acids in length containing a protease cleavage site. The donor
 CC and acceptor moieties exhibit fluorescent resonance energy transfer
 CC (FRET) when the donor moiety is excited. The constructs are used in
 CC enzymatic assays and can be used to isolate new enzymes or enzyme
 CC inhibitors or promoters. The specific activity of enzyme (in vivo and in
 CC vitro) and compounds altering enzyme activity can be obtained. FRET and
 CC hence activity of specific compounds is measured from the acceptor or
 CC donor moiety or maybe obtained using a ratio between the two.

XX Sequence 238 AA:

Query Match 98.6%; Score 1258; DB 18; Length 238;
 Best Local Similarity 98.7%; Pred. No. 2e-123;
 Matches 235; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 60
 DE 1 MSKGEELFTGVVPLVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 60
 QY 62 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 DE 62 VTITSYGVQCFSRYPDHHKRDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKEGDTLV 120
 QY 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVNFKIRHNIEDGVVQGLAD 180
 DE 121 NRIELKGIDFEDGNILGHKLEYVNSHNYIMADKKNGIKVNFKIRHNIEDGVVQGLAD 180
 QY 181 HYQONTPTGQGVVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 238
 DE 181 HYQONTPTGQGVVLELDGVNCHKFSVSGEGSDATYGLTKLTKFCTTCKLVPWPPTL 238

RESULT 6
 AAW76371 standard: protein; 238 AA.

XX AAW76371;
 AC AAW76371;
 DI 03-DEC-1998 (first entry)
 XX A. victoria green fluorescent protein;
 XX Green fluorescent protein; GFP; jellyfish; detection: mutagenic activity;
 KW distribution;
 XX Aequorea victoria;
 XX J010234382-A;
 XX 08-SEP-1998;
 XX 27-FEB-1997; 97JP-0062370;
 XX 27-FEB-1997; 97JP-0062370;
 XX (DNA-) DINABEKKE KENYUSHO KK;
 PA WP1: 1998-515036/46;
 DR N-PSDB; AAW61751.

XX Green fluorescent protein - useful for determining promoter and
 PT targeting activity in a cell;
 XX Disclosure: Page 6-7; 11pp; Japanese.

XX This sequence represents a wild-type green fluorescent protein, GFP,
 CC isolated from Aequorea victoria. This protein is used in the
 CC construction of a mutant protein which emits fluorescence 6 times
 CC higher than the conventional fluorescent protein. The nuclear acid can
 CC be used for the recombinant production of the fluorescent protein and
 CC also for the determination of activity of a promoter to be tested,
 CC comprising steps of introducing the above vector to a host cell and

CC detecting the fluorescence emitted from the cell, they can additionally
 CC be used for detecting the targeting activity in the cell of an amino
 CC acid sequence to be tested in which the above fluorescent protein is
 CC introduced to a cell, and the distribution of the fluorescent protein in
 CC the cell is observed.

XX Sequence 238 AA:
 Query Match: 98.64; Score: 1254; Db: 124; Length: 238;
 Best Local Similarity: 96.74; Pred. No. 2e-123;
 Matches: 235; Conservative: 0; Mismatches: 3; Indels: 0; Gaps: 0;
 QY 1 MSKGEELFTGVPIVLELGDVNGHKFSVSGEGDATYKTLKFTTTSKLFVPEWTL 60
 DB 1 MSKGEELFTGVPIVLELGDVNGHKFSVSGEGDATYKTLKFTTTSKLFVPEWTL 60
 QY 61 VTLSYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKDCGNYKTRAEVKEGTLV 120
 DB 61 VTLSYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKDCGNYKTRAEVKEGTLV 120
 QY 121 NRLEKGIIDKEKDNILGHKLEYNNSHNHYIMADKGNKATKVNFKIRHNEATVGLAD 180
 DB 121 NRLEKGIIDKEKDNILGHKLEYNNSHNHYIMADKGNKATKVNFKIRHNEATVGLAD 180
 QY 181 HYQONTPTGSGVLLPNHNYSTOSALSKEPNEDRHVLLGFTVIAAGITRSGHELYK 238
 DB 181 HYQONTPTGSGVLLPNHNYSTOSALSKEPNEDRHVLLGFTVIAAGITRSGHELYK 238

RESULT 7
 AAW76105
 ID AAW76105 standard; Protein: 238 AA.
 AC AAW76105;
 XX
 XX 18-NOV-1998 (first entry)
 DE A. victoria green fluorescent protein.
 KW Green fluorescent protein; GFP; mutant; jellyfish excitation;
 KW chromophore.
 XX
 CS Aequorea victoria.
 XX
 PN US804387-A. 9705-070132.
 XX
 XX 31-JAN-1997; 9705-070132.
 XX
 XX 01-FEB-1998; 9605-001096.
 XX 31-JAN-1997; 9705-070132.
 XX (STR) UNIV LELAND STANFORD JUNIOR.
 XX
 PI Cornack BP, Falkow S, Vardivia RH;
 XX
 XX WPI: 1998-505643/42.
 XX
 XX DNA encoding mutant green fluorescent protein. With
 XX greater fluorescence intensity than wild type proteins, useful for
 XX studying gene expression and protein localization.
 XX
 PS Disclosure: Page 1; 15pp; English.

XX
 CC This sequence represents the green fluorescent protein (GFP) from
 CC Aequorea victoria. This sequence is used in a method to discover GFP's
 CC mutated in the chromophore region that fluoresce more brightly than
 CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 CC method for analyzing a cell containing the GFP. The GFP can also be fused
 CC to a protein and used to identify the intracellular localization of a
 CC protein of interest. A regulatory element could be operatively connected
 CC to a coding portion encoding a mutant GFP and exposed to an environmental

CC stimulus. The fluorescence signal from the cell then measures
 CC the effect of the stimulus on the regulatory element.
 CC Note: This sequence does not appear in the specification but has
 CC been retrieved from the EMBL record contained in W05104.

XX Sequence 238 AA:
 Query Match: 98.64; Score: 1254; Db: 19; Length: 238;
 Best Local Similarity: 96.74; Pred. No. 2e-123;
 Matches: 235; Conservative: 0; Mismatches: 3; Indels: 0; Gaps: 0;
 QY 1 MSKGEELFTGVPIVLELGDVNGHKFSVSGEGDATYKTLKFTTTSKLFVPEWTL 60
 DB 1 MSKGEELFTGVPIVLELGDVNGHKFSVSGEGDATYKTLKFTTTSKLFVPEWTL 60
 QY 61 VTLSYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKDCGNYKTRAEVKEGTLV 120
 DB 61 VTLSYGVGCFSRYPDHNKREDFKSAKPGYVCEITFFKDCGNYKTRAEVKEGTLV 120
 QY 121 NRLEKGIIDKEKDNILGHKLEYNNSHNHYIMADKGNKATKVNFKIRHNEATVGLAD 180
 DB 121 NRLEKGIIDKEKDNILGHKLEYNNSHNHYIMADKGNKATKVNFKIRHNEATVGLAD 180
 QY 181 HYQONTPTGSGVLLPNHNYSTOSALSKEPNEDRHVLLGFTVIAAGITRSGHELYK 238
 DB 181 HYQONTPTGSGVLLPNHNYSTOSALSKEPNEDRHVLLGFTVIAAGITRSGHELYK 238

RESULT 8
 AAW40479
 ID AAW40479 standard; Protein: 238 AA.
 AC AAW40479;
 XX
 XX 15-SEP-1998 (first entry)
 DE A. victoria green fluorescent protein.
 KW Green fluorescent protein; protein kinase; phosphate donor;
 KW phosphorylation; drug screening; receptor-ligand binding; signaling;
 KW protein-protein interaction; kinase activation.
 XX
 CS Aequorea victoria.
 XX
 XX Key: Location/Qualifiers
 XX Protein: 1..238
 XX /cote= "partial sequence"
 XX
 XX W09632571-A1.
 XX
 XX 22-JAN-1998.
 XX
 XX 16-JUL-1997; 9705-0512410.
 XX
 XX 16-JUL-1998; 9805-0640873.
 XX 16-JUL-1998; 9805-0679865.
 XX 16-JUL-1998; 9805-0690876.
 XX
 XX (PDB) UNIV CALIFORNIA.
 XX
 XX Quidt A5, Tsien RY;
 XX
 XX WPI: 1998-110416/13
 XX N-PSDB: AAW11125.
 XX
 XX Assays for protein kinase and modulators - using a fluorescent
 XX protein substrate which exhibits a different fluorescent property in
 XX the phosphorylated and in-phosphorylated state
 XX
 PS Disclosure: Fig 1; 65pp; English.
 XX
 CC This sequence represents a green fluorescent protein from Aequorea
 CC victoria which is used in a method to determine if a sample contains

DB 61 VTTSSVGVQCFSRYPDHMKRDFKSAHSGVYQVQKTIFFKIDGNYKTRAEVKEFEDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKOKNGIKVNFKIRHNIEGGVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKOKNGIKVNFKIRHNIEGGVQLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTOSALSQDPNKRKHVMVLLGFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTOSALSQDPNKRKHVMVLLGFVTAAGITHGMDELYK 238

RESULT 12
 ID AAW22101 standard; protein: 238 AA.
 AC AAW22101:
 XX 10-MAR-1998 (first entry)
 XX Aequorea victoria protein GFP10 mutant F66L-S65L.
 KW Green fluorescent protein; GFP; jellyfish; GFP10 gene; reporter;
 KW humanise; expression; mutant.
 XX Aequorea victoria.
 OS
 XX KEY Location/Qualifiers
 FE Misc-difference 64 /Label: F66L
 FI /note: "Wild-type Phe is replaced by Leu"
 FT Misc-difference 65 /Label: S65L
 FE /note: "Wild-type Ser is replaced by Thr"
 FT W09726333 ALA
 XX 24-JUL-1997
 XX 17-JAN-1997: 9TWC-00755.
 XX 18-JAN-1996: 9GJS-0586201.
 XX (UFL) UNIV FLORIDA RES FOUND INC.
 XX Hauswirth WW, Rzyzcka N, Wolotukhin G;
 XX W21: 1997-383337/45.
 XX Humanised green fluorescent protein gene - optimised to provide high
 PT level expression in mammalian cells, used e.g. to label or identify
 FT cells, to locate proteins, etc.
 XX
 XX Claim 115: Page 7: 15pp; English.
 XX
 CC This sequence represents a mutant form of the green fluorescent protein
 CC GFP10 i.e. F66L-S65L. Green fluorescent proteins (GFP's) could be used
 CC as reporter molecules as they absorb blue light and emit green light
 CC without requiring any cofactors, substrates, or additional gene products
 CC allowing GFP detection in living cells provided meaningful gene
 CC expression is achieved. By providing humanised GFP, adapted for
 CC expression in mammalian and human cells, problems associated with
 CC wild-type jellyfish GFP e.g. variable and low expression levels should be
 CC overcome.
 CC Note: The present sequence does not appear in the specification. It
 CC has been made by modifying the wild-type GFP10 amino acid sequence found
 CC in AAW1295.
 XX
 XX Sequence 238 AA:
 SQ

QY 1 MSKGEELFTGWPIILVLDGNGHGHKPSVSGEGEDATYCKLTLKFLCTGKLPVWPTL 60
 DB 1 MSKGEELFTGWPIILVLDGNGHGHKPSVSGEGEDATYCKLTLKFLCTGKLPVWPTL 60
 QY 61 VTTSSVGVQCFSRYPDHMKRDFKSAHSGVYQVQKTIFFKIDGNYKTRAEVKEFEDTLV 120
 DB 61 VTTSSVGVQCFSRYPDHMKRDFKSAHSGVYQVQKTIFFKIDGNYKTRAEVKEFEDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKOKNGIKVNFKIRHNIEGGVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNYNHSHVYIMADKOKNGIKVNFKIRHNIEGGVQLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTOSALSQDPNKRKHVMVLLGFVTAAGITHGMDELYK 238
 DB 181 HYQONTPIGDPVLLPDNHYLSTOSALSQDPNKRKHVMVLLGFVTAAGITHGMDELYK 238

RESULT 13
 ID AAW76109 standard; protein: 238 AA.
 AC AAW76109:
 XX 18-NOV-1998 (first entry)
 XX A. victoria green fluorescent protein mutant S72A.
 KW Green fluorescent protein; GFP; mutant; jellyfish; excitation;
 KW chromophore.
 OS Aequorea victoria.
 XX KEY Location/Qualifiers
 FE Misc-difference 72 /Label: S72A
 FT /note: "Wild-type Ser is replaced by Ala"
 XX US5404367-A.
 XX US-SEP-1998.
 XX 31-JAN-1997: 97JS-0791342.
 XX 31-SEP-1996: 96JS-0070960.
 XX 31-JAN-1997: 97JS-0791342.
 XX (SIR) UNIV LEIAND STANFORD JUNIOR.
 XX Cornack BP, Falkow S, Valdivia PH;
 XX W21: 1998-505643/43.
 XX
 XX DNA encoding mutant green fluorescent pigment proteins - with
 XX greater fluorescence intensity than wild-type proteins, useful for
 XX studying gene expression and protein localisation
 XX Claim 5: Page 7: 15pp; English.
 XX This sequence represents a mutant green fluorescent protein (GFP) from
 XX Aequorea victoria in which a Ser residue at position 72 in the wild
 XX type protein is replaced by an Ala. These mutations occur in the
 XX chromophore region. GFP mutants are used in a method to discover GFP's
 XX mutated in the chromophore region that fluoresce more brightly than
 XX wild-type GFP upon excitation at 488 nm. These mutants can be used in a
 XX method for analysing a cell containing the GFP. The GFP can also be
 XX fused to a protein and used to identify the intracellular localisation of
 XX a protein of interest. A regulatory element could be operative, if
 XX connected to a coding portion encoding a mutant GFP and exposed to an
 XX environmental stimulus. The fluorescence signal from the cell
 XX then measures the effect of the stimulus on the regulatory element.
 XX NOTE: This sequence does not appear in the specification but has been
 XX constructed from the wild-type GFP protein represented in AAW76106

```

XX SQ Sequence 235 AA;
Query Match 98.4%; Score 1255; Pos 19; Length 235;
Best Local Similarity 98.3%; Pred. No. 4 (10-12);
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSKGEELPTGVVFLVILVDELGVNCHKFSVSQSGHSDATYKGLIKFKFTCTGKLVVWPTL 60
DB 1 MSKGEELPTGVVFLVILVDELGVNCHKFSVSQSGHSDATYKGLIKFKFTCTGKLVVWPTL 60
QY 61 VTTLSYVQCFPSRYPOHKKHDFKFSAMPEYVOERTIFFKDDGNYKTRAPKPESTLV 120
DB 61 VTTLSYVQCFPSRYPOHKKHDFKFSAMPEYVOERTIFFKDDGNYKTRAPKPESTLV 120
QY 121 NRLEKGSDFKEDGNTLGHKLEYNNSHNVYIMADKSKNGIKVNFKIRHNTEDSGVGLAD 180
DB 121 NRLEKGSDFKEDGNTLGHKLEYNNSHNVYIMADKSKNGIKVNFKIRHNTEDSGVGLAD 180
QY 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHMYLGLPVTAARITHGHEDEYK 240
DB 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHMYLGLPVTAARITHGHEDEYK 240
RESULT 14
AAW76110 standard: protein: 235 AA.
XX ID AAW76110;
XX AC AAW76110;
XX DT 18-NOV-1998 (first entry)
XX DE A. victoria green fluorescent protein mutant: S65A.
XX KW Green fluorescent protein: GFP; mutant: jellyfish; excitation;
XX OS Aequorea victoria.
XX FS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 65 /label- S65A
FT /note- "Wild-type Ser is replaced by Ala"
XX PN US5564387-A.
XX PC 08-SEP-1994.
XX XX 31-JAN-1997; 97US-0791332.
XX PR 01-FEB-1996; 96US-0610960.
XX PR 31-JAN-1997; 97US-0751332.
XX PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX PI Cormack RP, Falkow S, Valdivia RH.
XX WP1: 1998-505643/43.
XX FS Claim 6; Page 1; 15pp; English.
XX CC This sequence represents a mutant green fluorescent protein (GFP) from
CC Aequorea victoria in which a Ser residue at position 65 in the wild
CC type protein is replaced by an Ala. These mutations occur in the
CC chromophore region. GFP mutants are used in a method to discover GFP's
CC mutated in the chromophore region that fluoresce more brightly than
CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
CC method for analyzing a cell containing the GFP. The GFP can also be

```

```

CC fused to a protein and used to identify the intracellular localisation of
CC a protein of interest. A regulatory element could be operatively
CC connected to a coding portion encoding a mutant GFP and exposed to an
CC environmental stimulus. The fluorescence signal from the cell
CC then measures the effect of the stimulus on the regulatory element.
CC NOTE: This sequence does not appear in the specification but has been
CC constructed from the wild-type GFP protein represented in AAW71106
XX Sequence 238 AA;
Query Match 98.4%; Score 1255; Pos 19; Length 238;
Best Local Similarity 98.3%; Pred. No. 4 (10-12);
Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSKGEELPTGVVFLVILVDELGVNCHKFSVSQSGHSDATYKGLIKFKFTCTGKLVVWPTL 60
DB 1 MSKGEELPTGVVFLVILVDELGVNCHKFSVSQSGHSDATYKGLIKFKFTCTGKLVVWPTL 60
QY 61 VTTLSYVQCFPSRYPOHKKHDFKFSAMPEYVOERTIFFKDDGNYKTRAPKPESTLV 120
DB 61 VTTLSYVQCFPSRYPOHKKHDFKFSAMPEYVOERTIFFKDDGNYKTRAPKPESTLV 120
QY 121 NRLEKGSDFKEDGNTLGHKLEYNNSHNVYIMADKSKNGIKVNFKIRHNTEDSGVGLAD 180
DB 121 NRLEKGSDFKEDGNTLGHKLEYNNSHNVYIMADKSKNGIKVNFKIRHNTEDSGVGLAD 180
QY 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHMYLGLPVTAARITHGHEDEYK 248
DB 181 HYQONTPTGDEGVILPQNHLYSTQSLSKDNEKRDHMYLGLPVTAARITHGHEDEYK 248
RESULT 15
AAW76111 standard: protein: 238 AA.
XX ID AAW76111;
XX AC AAW76111;
XX DT 18-NOV-1998 (first entry)
XX DE A. victoria green fluorescent protein mutant V68L.
XX KW Green fluorescent protein: GFP; mutant: jellyfish; excitation;
XX OS Aequorea victoria.
XX FS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 68 /label- V68L
FT /note- "Wild-type Val is replaced by Leu"
XX PN US5564387-A.
XX PC 08-SEP-1994.
XX XX 31-JAN-1997; 97US-0791332.
XX PR 01-FEB-1996; 96US-0610960.
XX PR 31-JAN-1997; 97US-0751332.
XX PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX PI Cormack RP, Falkow S, Valdivia RH.
XX WP1: 1998-505643/43.
XX FS Claim 6; Page 1; 15pp; English.
XX CC This sequence represents a mutant green fluorescent protein (GFP) from
CC Aequorea victoria in which a Val residue at position 68 in the wild
CC type protein is replaced by a Leu. These mutations occur in the
CC chromophore region. GFP mutants are used in a method to discover GFP's
CC mutated in the chromophore region that fluoresce more brightly than
CC wild-type GFP upon excitation at 488 nm. These mutants can be used in a
CC method for analyzing a cell containing the GFP. The GFP can also be

```

